

P25921

10/506610
DT15 Rec'd PCT/PTO 10 SEP 2004

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : N. TOMIOKA et al.

Appl. No: : Not Yet Assigned

PCT Branch

Filed : Concurrently Herewith

PCT/JP03/02847

For : METHOD OF GENERATING MOLECULE-FUNCTION NETWORK


CLAIM OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Applicant hereby claims the right of priority granted pursuant to 35 U.S.C. 119 based upon Japanese Application No. 2002-64994 filed 11 March 2002. The International Bureau already should have sent a certified copy of the Japanese application to the United States designated office. If the certified copy has not arrived, please contact the undersigned.

Respectfully submitted,
N. TOMIOKA et al.


Bruce H. Bernstein
Reg. No. 29,027
33,329

September 10, 2004
GREENBLUM & BERNSTEIN, P.L.C.
1950 Roland Clarke Place
Reston, VA 20191
(703) 716-1191

REC'D PCT/PTO 10 SEP 2004

#2
PCT/JP 03/02847

日 本 国 特 許 庁

JAPAN PATENT OFFICE

10/506610

11.03.03

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 3月11日

出 願 番 号

Application Number:

特願2002-064994

[ST.10/C]:

[JP2002-064994]

出 願 人

Applicant(s):

株式会社医薬分子設計研究所

REC'D 05 MAY 2003

WIPO

PCT

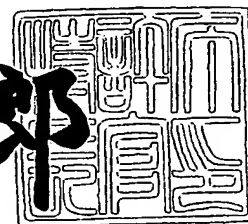
**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

2003年 4月15日

特許庁長官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3026514

【書類名】 特許願

【整理番号】 A21157M

【提出日】 平成14年 3月11日

【あて先】 特許庁長官 殿

【発明者】

 【住所又は居所】 東京都文京区本郷 5 - 1 6 - 6

 【氏名】 板井 昭子

【特許出願人】

 【識別番号】 597051148

 【氏名又は名称】 株式会社医薬分子設計研究所

【代理人】

 【識別番号】 100096219

 【弁理士】

 【氏名又は名称】 今村 正純

【選任した代理人】

 【識別番号】 100092635

 【弁理士】

 【氏名又は名称】 塩澤 寿夫

【選任した代理人】

 【識別番号】 100095843

 【弁理士】

 【氏名又は名称】 釜田 淳爾

【手数料の表示】

 【予納台帳番号】 038357

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

【包括委任状番号】 9707349

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 分子機能ネットワークの生成方法

【特許請求の範囲】

【請求項 1】 修飾状態、活性／不活性状態、集合状態、及び構造変化からなる群から選ばれる項目を含む 1 又は 2 以上の項目により生体分子の情報を階層化して保存したデータベースを用いてコネクト検索を行うことを特徴とする、分子機能ネットワークの生成方法。

【請求項 2】 生体分子対の情報に該生体分子対が形成される条件を含む生体分子連鎖データベースを用いてコネクト検索を行うことを特徴とする、分子機能ネットワークの生成方法。

【請求項 3】 疾患に関する情報をグループ化及び／又は階層化されたデータ項目として保存し、該データ項目間の関連付け情報を保存した病態連鎖データベースを用いてコネクト検索を行うことを特徴とする、分子機能ネットワークの生成方法。

【請求項 4】 生体分子連鎖データベースを用いて、信頼性コード、作用臓器、及び生体分子対の方向性からなる群から選ばれるデータ項目を含む 1 又は 2 以上のデータ項目に条件を設定することにより、生体分子対にフィルターをかけてコネクト検索を行って分子機能ネットワークを生成する方法。

【請求項 5】 生体分子連鎖データベースを用いてコネクト検索により生成した分子機能ネットワークを、信頼性コード、作用臓器、及び生体分子対の方向性からなる群から選ばれるデータ項目を含む 1 又は 2 以上のデータ項目に基づいてスコア付けすることにより、該分子機能ネットワークを絞り込む方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、生体イベントの情報を含む生体分子データベースの作成方法と利用方法に関する。

【0002】

【従来技術】

生体中には、DNA、RNA、蛋白質、多糖類などの生体高分子の他に、アミノ酸、核酸、脂質、糖質、一般低分子化合物など多様な分子が存在して、それぞれの機能を担っている。生体系の特徴は、単に多様な生体分子から構成されているというだけでなく、機能の発現をはじめとする生体内のすべての現象が生体分子間の特異的な結合を通じて起きることにある。この特異的な結合では、共有結合は形成されずに、分子間力のみによって安定化された複合体が形成される。従って、生体分子はそれが単独で存在する状態と複合体状態の間で、平衡状態で存在することになるが、特定の生体分子間では複合体状態の安定化が大きいために、この平衡が複合体側に著しく偏っている。その結果、多数の他分子の存在下で、かなり希薄な濃度でも実質的に特定の相手を識別して結合することができる。酵素反応においては、基質は酵素との複合体を形成した状態で所定の化学変換を受けた後に反応生成物として放出されるし、シグナル伝達においては、メディエーター分子の標的分子への結合に伴う標的分子の構造変化を通じて細胞外のシグナルが細胞内に伝達される。

【0003】

近年、ゲノム研究の進展は目覚ましく、ヒトをはじめとする多様な生物種についてゲノム配列が解明され、遺伝子や遺伝子産物である蛋白質の配列、臓器毎の蛋白質の発現、蛋白-蛋白相互作用等についてもゲノムワイドな体系的な研究が進みつつある。それらの研究成果の大部分はデータベースとして公開されて、世界中の利用に供されている。遺伝子や蛋白質の機能、疾患の原因又は背景となる遺伝子の推定、遺伝子多型との関わりについても少しずつ解明が進み、遺伝情報に基づく医療や新薬開発への期待が高まってきた。

【0004】

一方、遺伝情報の担い手は核酸であるが、エネルギー代謝、物質変換、シグナル伝達など生命機能の殆どは、核酸以外の分子が担っている。蛋白質は遺伝子という設計図に基づいて直接生産される点で他のカテゴリーの分子と異なり、その種類も多い。酵素も、低分子性の生体内生理活性化合物の標的分子も、蛋白性の生体内生理活性化合物の標的分子も（糖で修飾されていることも多いが）、すべて蛋白質である。疾患の根本原因はともかくとして、多くの疾患や症状が

蛋白質や低分子化合物の量やバランス、場合によっては質（機能）の異常の結果であると考えられている。既存医薬の殆どが、蛋白質を標的としてその機能を制御する化合物である。蛋白質と違い、核酸の立体構造が低分子医薬の標的として特異性を発揮しにくいという理由もあるが、抗生物質や抗菌剤、或いは農薬の殺虫剤や抗カビ剤に至るまで、標的は蛋白質である。

【0005】

従って、遺伝情報に基づいた医療や新薬開発を行うには、生体内のそれぞれの蛋白質と低分子化合物の機能とそれらの分子間の特異的關係を明らかにする必要がある。さらに、異なる酵素が次々に関与して必要な分子が生合成されることや、異なる分子が次々に結合することでシグナルが伝達されることなどから、これらの分子は機能上や生合成上で互いに直接又は間接のつながりをもっており、そのつながり（分子機能ネットワーク）の情報が重要である。さらに、これまでの研究で、メディエーターやホルモンなど多彩な臨床症状や生理現象や生体反応の発現に直接関わっている分子が多数明らかにされており、適切な治療には分子機能ネットワークとの関係付けが不可欠である。また、創薬戦略において、副作用リスクを考慮しつつ適切な新薬開発の標的分子を設定するためには、標的分子を含む分子機能ネットワークの考慮が必要である。

【0006】

蛋白質に関連したデータベースとしては、SwissProt (the Swiss Institute of Bioinformatics (SIB), Europaen Bioinformatics Institute (EBI))、PIR (National Biomedical Research foundation (NBRF))があり、いずれも配列情報の他、生物種、機能、機能メカニズム、発見者、文献その他のアノテーションの情報をもっている。

【0007】

分子のつながりに注目した分子ネットワーク型データベースのうち、代謝経路に関するものとしては、KEGG (金久ら、京都大学)、Biochemical Pathways (Boehringer Mannheim), WIT (Russian Academy of Sciences), Biofrontier (呉羽化学)、Protein Pathway (AxCell), bioSCOUT (LION), EcoCyc (DoubleTwist), UM-BBD (Minnesota Univ.) がある。

【0008】

KEGGのPATHWAYデータベースにはメタボリックパスウェイとシグナルトランスダクションパスウェイがあり、前者は物質代謝、エネルギー代謝に関わる一般低分子化合物の代謝経路を、後者はシグナル伝達系の蛋白質を扱っている。いずれも予め定義された分子ネットワークを静的なGifファイルで提供している。前者では、酵素とリガンドの情報を別のテキスト形式の分子データベースLIGAND（金久ら、京都大学）及びENZYME（IUPAC-IUBMB）から取り込むようになっている。生理活性ペプチドの生成に関わる酵素や標的生体分子は含まれない。

【0009】

EcoCycは大腸菌の物質代謝をデータベース化したもので、個々の酵素反応に関するデータと、既知のパスウェイに関するデータ（該パスウェイに属する酵素反応の集合として表される）に基づいて、パスウェイの図式的な表示を行なっている。EcoCycの検索機能としては、分子名やパスウェイ名に対する文字列や略号による検索法が提供されているが、任意の分子を指定して新たなパスウェイを検索することはできない。

【0010】

シグナル伝達に関するものとしては、CSNDB（国立医薬品食品衛生研究所、日本）、SPAD（久原ら、九州大学）、Gene Net（Institute of Cytology & Genetics Novosibirsk, Russia）、GeNet（Maria G. Samsonova）がある。蛋白-蛋白相互作用のデータベースとしては、DIP（UCLA）、PathCalling（CuraGen）、ProNet（Myriad）がある。遺伝子・蛋白発現のデータベースとしては、BodyMap（東京大学、大阪大学）、SWISS-2DPAGE（Swiss Institute of Bioinformatics）、Human and mouse 2D PAGE database（Danish Centre for Human Genome Research）、HEART-2DPAGE（GermanHeart）、PDD Protein Disease Databases（NIMH-NCI）、Washington University Inner Ear Protein Database（Washington Univ.）、PMMA-2DPAGE（Purkyne Military Medical Academy）、Mito-Pick（CEA, France）、Molecular Anatomy Laboratory（Indiana University）、Human Colon Carcinoma Protein Database（Ludwig Institute for Cancer Research）がある。生体反応シミュレーション型の分子ネットワークとしてはE-Cell（富田ら、慶応大学）、e E.

coli (B. Palsson), Cell (D. Lauffenburger, MIT), Virtual Cell (L. Leow, Conneticut Univ.), Virtual Patient (Entelos, Inc.) がある。

【0011】

生体分子と機能の関係については、広範な蛋白の情報を集めたSwissProt の他、COPE (University of Munich) がサイトカインの機能情報をテキスト形式で公開している。ARIS ((株)日本電子計算)は、医学・薬学分野を中心に約400の国内雑誌、20の海外雑誌から医薬品の副作用、相互作用、農薬・化学物質による中毒に関わる文献情報を収録しているが、生体分子の生理的作用や細胞レベル以上での応答についてのデータベースはこれまでにない。遺伝子と疾患については、遺伝病と蛋白質のアミノ酸変異の情報を集めたOMIM (NIH) がある。いずれもテキスト形式のデータが記述されていて、キーワードで検索できる。

【0012】

分子間のつながりに注目した既存データベースの問題点は以下の点にある。分子ネットワーク型データベースは含まれる分子と分子間のつながりが分かっている系に対して作られており、予め分子間の関係を考慮して配置することが可能なために、静的なGif 等での表現で済んできた。しかし、この方式では、新しい分子や分子間のつながりの追加が困難である。今後存在が明らかにされるであろう分子も含めると10万以上あり(KEGGで扱っている分子数は医薬分子を含めて約1万)、さらにそれらの分子間の関わりが今後の研究により解明されると、分子ネットワークの複雑さは加速度的に増すことが予想される。新しい分子の追加に対応でき、膨大な分子と分子間の関係の情報を保持しながら、必要な情報を含む部分的分子ネットワークを生成できる新しい方法が必要である。

【0013】

2001年9月7日現在、KEGGでは分子間のつながりを2分子の対の情報として保存し、この情報を用いてメタボリックパスウェイ中の任意の2分子間をつなぐ経路を探索することが可能になっている。しかし、このような経路探索問題には、2分子間をつなぐ経路が長くなるにつれて計算時間が級数的に増加するという問題がある。

【0014】

一方、テキスト形式のデータベースでは分子データの追加に限界はない。しかし、個々の分子のデータから機能上又は生合成上関連ある分子を次々に探索を繰り返して、多数の分子のつながりを示す分子ネットワークを生成するには困難が伴う。検索時に必要な分子についてのつながりが動的にかつ自動的に得られるようなデータの保持方法と検索方法を開発する必要がある。また、疾患や病態を分子レベルで理解するためには、生体分子やそのネットワークと生体の応答や生理作用との関係を記述する新しい工夫が必要である。

【 0 0 1 5 】

【発明が解決しようとする課題】

本発明の課題は、多様な生体応答や現象を生体分子の機能とその分子間の関わりにおいて理解するための仕組みと方法を提供することにある、より具体的には生体分子の情報を生体応答と結びつけることのできるデータベースとその検索法を提供することにある。さらには、それらの膨大な情報から、任意の生体応答又は生体分子に関連するシグナル伝達経路や生合成経路のみを迅速に効率よく取り出し、有望な創薬ターゲットや副作用リスクを推定するための方法を提供することも本発明の課題の1つである。

【 0 0 1 6 】

【課題を解決するための手段】

本発明者らは、上記の課題を解決すべく鋭意努力した結果、直接結合する生体分子の対をパーツとした情報を集積することで生体分子間のつながりを網羅し、生体分子の中で生体応答の発現に直接関わるキイ分子については標的生体分子との対の情報に生理作用、生体応答、臨床症状等からなる生体イベントの情報を加えて保存し、指定した1以上の任意の生体分子又は生体イベントを含むつながりを自動的に次々と探索して分子機能ネットワークを作成することによって上記の課題を解決できることを見出し、それらの発明について特許出願した（PCT/JPO1/07830号明細書）。

【 0 0 1 7 】

すなわち、直接結合する生体分子対に関する情報を蓄積した生体分子連鎖データベースを用いて分子機能ネットワークを生成する方法がPCT/JPO1/07

830号明細書に記載されている。この発明の好ましい態様によれば、生体イベントの情報をもつ生体分子連鎖データベースを用いて生体イベントの情報と関係付けた分子機能ネットワークを生成する上記の方法；生体分子自身に関する情報を含む生体分子情報データベースを用いる上記の方法；及び生体イベントの情報と関係付けた医薬分子を含む分子機能ネットワークを生成する上記の方法が提供されている。また、直接結合する生体分子対に関する生体イベントの情報を蓄積した生体分子連鎖データベースを用いて任意の生体分子又は医薬分子が直接又は間接に関わる生体イベントを推定する方法が提供された。さらに、生体分子が、外部のデータベース又は文献中の遺伝子でコードされる蛋白質である場合に、該生体分子の分子略号と該遺伝子の名称又はID又は略称を対応付けるデータベースを作成することにより、遺伝子の多型や発現の情報を分子機能ネットワークを用いて解析する方法が提供されている。

【0018】

上記発明のさらに好ましい態様により、ネットワーク上での連なりに基づいてグループ化した生体分子対データをサブネットとして、帰属するサブネット及びサブネット間の包含関係に基づいて、分子機能ネットワークを階層化することを特徴とする上記の方法；生体分子対に関する情報を該分子対の帰属経路名・帰属サブネット名等に基づいて階層化して保存することを特徴とする上記の方法；生体分子自身に関する情報を遺伝子群からの発現パターンや細胞表面への出現パターン等に基づいて階層化して保存することを特徴とする上記の方法；生体イベントに関する情報を該イベントの上位概念による分類や病態イベントとの関連に基づいて階層化して保存することを特徴とする上記の方法が提供されている。さらに、生体分子対に関する上位階層、生体分子自身に関する上位階層、生体イベントに関する上位階層のそれぞれの保存項目に対して、項目間の関連性や依存関係等に関する情報を保存することを特徴とする上記の方法；生体分子情報データベース又は生体分子連鎖データベースに保存された階層化の情報を利用することにより、分子機能ネットワークの生成を容易にすることを特徴とする上記の方法；生体分子情報データベース又は生体分子連鎖データベースに保存された階層化の情報を利用することにより分子機能ネットワークの表現での詳細さを制御すること

を特徴とする上記の方法も上記発明により提供されている。

【0019】

本発明者らはさらに鋭意研究を行った結果、以下に示す方法及びデータベースを提供することに成功した。

1. 生体分子連鎖データベースを用いて、信頼性コード、作用臓器、生体分子対の方向性その他のデータ項目の1又は2以上の組み合わせにより生体分子対にフィルターをかけてコネクト検索を行い生成した分子機能ネットワーク又は分子機能ネットワークを生成する方法。
2. 生体分子連鎖データベースを用いてコネクト検索により生成した分子機能ネットワークを、信頼性コード、作用臓器、生体分子対の方向性その他のデータ項目の1又は2以上の組み合わせによりスコア付けすることにより、該分子機能ネットワークを絞り込む方法。
3. 端末で閲覧中の情報のなかに、生体分子情報データベース、生体分子連鎖データベース、医薬分子情報データベース、医薬分子連鎖データベース及び病態連鎖データベースのいずれかに含まれるデータ項目に該当する項目が存在した場合に、該項目を強調して表示する方法。
4. 端末で閲覧中の情報のなかに、コネクト検索により生成した分子機能ネットワーク中のいずれかのデータ項目に該当する項目が存在した場合に、該項目を強調して表示する方法。
5. 端末で閲覧中の情報のなかに、生体分子情報データベース、生体分子連鎖データベース、医薬分子情報データベース、医薬分子連鎖データベース及び病態連鎖データベースのいずれかに含まれるデータ項目に該当する項目が存在した場合に、該項目を含むデータベースの情報を表示する方法。
6. 端末で閲覧中の情報のなかに、コネクト検索により生成した分子機能ネットワーク中のいずれかのデータ項目に該当する項目が存在した場合に、該分子機能ネットワーク上で該項目を強調表示する方法。

【0020】

【発明の実施の形態】

本発明の理解のためにPCT/JPO1/07830号明細書の開示を参照する

ことは有用である。上記PCT/JPO1/07830号明細書の開示の全てを参照として本明細書の開示に含める。

本明細書における用語の意味又は定義は以下の通りである。

「生体」とは、例えばオルガネラ、細胞、組織、臓器、生物個体、又は集合体など、生物に寄生する生命体を含む概念とする。

「生体イベント」とは、生体において内因的に又は外因的に現れるすべての現象、応答、反応、症状を含む概念とする。具体的な例として、転写、細胞の遊走、細胞の接着、細胞分裂、神経回路興奮、血管収縮、血圧上昇、血糖低下、発熱、痙攣、異種生物及びウイルスなど寄生体による感染その他を挙げることができる。また、光や熱などの生体外部からの物理的な刺激に対する応答も生体イベントの概念に含めることができる。

【0021】

「病態イベント」とは「生体イベント」に含まれる概念であって、「生体イベント」が量的又は質的にある閾値を超え、疾患又は病態と判断できる状態をいう。例として、血圧上昇の「生体イベント」が異常に亢進した結果の「病態イベント」としては高血圧又は高血圧症、血糖が正常範囲に制御できなくなった「病態イベント」としては高血糖又は糖尿病を挙げることができる。また、上記の例のように単一の生体イベントに関連するものだけではなく、複数の種類の生体イベントに関連している病態イベントもある。

「生体分子」とは、生体中に存在する核酸、蛋白質、脂質、糖質、一般低分子化合物その他のあらゆる構造の有機分子及びその集合体を指し、金属イオン、水、プロトンを含んでいてもよい。

【0022】

「キイ分子」とは、生体分子のうち、主としてメディエーター、ホルモン、神経伝達物質、オータコイド等の分子群を指す。多くの場合、体内に特定の標的生体分子が存在して、その分子への直接結合が上記の「生体イベント」の引き金となることが知られている。これらの分子は生体内で生成されて作用を発現しているが、一般には生体系の外部から与えた場合にもその量に対応した生体イベントを発現する。具体的例として、アドレナリン、アンジオテンシンII、インシュリン

、エストロゲンその他を挙げることができる。

また標的生体分子が確認されていなかったり、標的生体分子への直接結合が確認されていなかったりする生体分子であっても、該生体分子の量的又は質的な変動が生体イベントの亢進・上昇又は低下・減少をもたらしたり、分子機能ネットワークやサブネットに変動を与えたりすることが分かっている場合には、該分子を「キイ分子」として扱ってもよい。

【0023】

「標的生体分子」とは、メディエーター、ホルモン、神経伝達物質、オータコイド等の生体分子又は医薬分子の受け皿となる特定の生体分子をいう。直接結合して特定のイベントを発現する。

「生体イベントのアップオアダウンの情報」とは、キイ分子又は標的生体分子の量的又は質的な変動に対応して生体イベントが亢進・上昇又は低下・減少する等の情報をいう。キイ分子の量が一定の閾値を超えた場合にはじめてその生体イベントが起きる場合も含む。

「分子略号」とは、分子名称の代わりに分子を識別又は指定する目的に付するもので、各分子にユニークに対応している必要がある。分子名を短縮した略号でも、分子名と無関係な英数字列でもよいが、短い文字列であることが望ましい。既に世界的に使用されている分子略号がある分子の場合には、それを用いるのが望ましい。1ヶの分子に対して異なる方式で付けられた複数の分子略号を与え、構造グループや機能などにより階層化することも可能である。

【0024】

「直接結合」とは、共有結合によらない分子間力によって安定な複合体を形成する、又は形成可能であることをいう。稀に共有結合が形成されることもあるが、この場合を含めた概念である。「相互作用」と呼ばれることも多いが、相互作用の方が広い意味を含んでいる。

「生体分子対」とは、生体中で直接結合できるか直接結合することが推定される生体分子の対をいう。具体的な例として、エストラジオールとエストロゲン受容体、アンジオテンシン変換酵素とアンジオテンシンⅠなどを挙げることができる。酵素反応における酵素と生成物の分子対の場合には、その複合体はあまり安定

とは言えないが、生体分子対に含めるものとする。また、例えばtwo-hybrid実験法で相互作用があるとされた2ヶの蛋白分子のように両者の相互の役割が明確でない場合を含めてもよい。光、音、温度変化、磁気、重力、圧力、振動などの生体外部からの物理的・化学的刺激についても、これらの刺激を仮想的な生体分子として扱い、対応する標的・生体分子との生体分子対を定義してもよい。

【0025】

「関連付け」とは、生体分子、サブネット、生体イベント、病態イベント、医薬分子、遺伝子、疾患に関するデータ項目の、いずれか2つのデータ項目の間に直接的又は間接的に関連性があると示すこと又は記録することをいう。「関連付け情報」とは「関連付け」することにより記録された情報のことをいう。

「構造コード」とは、生体分子がDNAか、RNAか、蛋白質か、ペプチドか、一般低分子か等の構造上の特徴を表す分類コードである。

「機能コード」とは、生体分子の分子レベルでの機能を表わす分類コードであり、例えば、「構造コード」が「蛋白質」である生体分子の場合は、膜受容体・核内受容体・トランスポーター・メディエーター・加水分解酵素・リン酸化酵素・脱リン酸化酵素等の分類を表わし、「構造コード」が低分子である生体分子の場合は、基質・生成物・前駆体・活性ペプチド・代謝産物等の分類を表わす。

【0026】

「関係コード」とは、生体分子対をなす2分子間の関係を表わす分類コードである。例えば、アゴニストと受容体であれば10、酵素と基質であれば21、基質と生成物であれば22というように類型化することができる。例えばtwo-hybrid実験法で相互作用があるとされた2ヶの蛋白分子のように2分子の相互の役割が明確でない場合は、その旨を区別したコードを用いることが望ましい。

「関係機能コード」とは、生体分子対をなす2分子の直接結合に伴う現象や変化を表わす分類コードであり、例えば加水分解、リン酸化、脱リン酸化、活性化、不活化等の分類を用いる。

「信頼性コード」とは、生体分子対毎に直接結合の信頼性のレベルや直接結合の根拠となった実験法などを示すためのコードである。

【0027】

「コネクト検索」とは、1以上の任意の生体分子又は生体イベントを指定して、それらを含む機能上又は生合成上関連した分子のつながりを自動的に探すことをいう。

コネクト検索において、信頼性コード又は作用臓器又は生体分子対の方向性その他の情報を1または2以上組み合わせることによって生体分子対にフィルターをかけながら実行することもできる。また、信頼性コード又は作用臓器又は生体分子対の方向性その他の情報を単独または複数組み合わせることによって探し出した分子のつながりにスコア付けを行って、望みの分子機能ネットワークを探し出すこともできる。

コネクト検索は、サブネット、病態イベント、医薬分子、遺伝子、疾患に関する情報などの任意の項目を指定して、生体分子又は生体イベントを指定した場合と同様に行うことができる。

【0028】

「分子機能ネットワーク」とは、生体分子連鎖データベースを用い、1以上の任意の生体分子又は生体イベントを指定してコネクト検索した結果得られる、機能上又は生合成上関連した分子のつながりをいう。

「医薬分子」とは、医薬として製造され治療に用いられる化合物の分子を指すが、医学・薬学研究用に用いられる化合物や特許明細書や文献に記載の化合物など生理活性が既知の化合物も含むものとする。

「生体イベントの情報と関係付ける」とは、ある生体イベントの発現に、ある生体分子又は医薬分子又は遺伝子情報又は分子機能ネットワークが関わっていることを示すか又は見つけることをいう。

「分子アノテーション」とは、データベース中の生体分子、医薬分子等の分子のデータに対して付加される情報のことである。

【0029】

「パターン化」とは、生体分子、生体分子対、生体イベント等に関する情報をデータベースに記録する際に、与えられた情報をそのまま記録するのではなく、予め決められたカテゴリーに該情報を分類し、該当するカテゴリーを表わす記号により該情報を表現することをいう。上記の「構造コード」、「機能コード」、「

関係コード」、「関係機能コード」について挙げられた例が、「パターン化」の例である。

「生成臓器」とは、生体分子が生成される臓器、組織、臓器又は組織内の部位、臓器又は組織内の特定の細胞、細胞内の部位などをいう。

「存在臓器」とは、生体分子が生成された後に蓄えられる臓器、組織、臓器又は組織内の部位、臓器又は組織内の特定の細胞、細胞内の部位などをいう。

「作用臓器」とは、生体分子又はキイ分子が生体イベントを発現する臓器、組織、臓器又は組織内の部位、臓器又は組織内の特定の細胞、細胞内の部位などをいう。

【 0 0 3 0 】

本発明の実施の一つの態様として図 1 に示す方法が提供される。まず、直接結合する 2 々の生体分子の対に関する情報を蓄積した「生体分子連鎖データベース」を作成する。生体分子の分子略号の付け方など、生体分子自身に関する情報をここに含めてもよいが、別データベースである「生体分子情報データベース」で行うのが望ましい。次に、上記の「生体分子連鎖データベース」から、1 以上の任意の分子を指定してコネクト検索を行ない、1 以上の生体分子の機能上又は生合成上のつながりの表現である「分子機能ネットワーク」を得る。

【 0 0 3 1 】

生体分子対のうち、少なくともキイ分子とその標的生体分子からなる生体分子対に対して生体イベントの情報を関連付けておくことにより、「分子機能ネットワーク」とともに、分子機能ネットワーク中の分子が直接又は間接に関わる生体イベントを推定することができる。さらに、キイ分子の量的又は質的な変動と生体イベントのアップオアダウンの関係の情報を付加しておくことにより、分子機能ネットワーク上の任意の分子の量的又は質的変動が生体イベントの亢進・上昇に働くか、抑制・低下に働くか等を推定できる。

【 0 0 3 2 】

「生体分子情報データベース」の主たる役割は、各生体分子の正式名称に対して分子略号又は ID を定義することであり、その他にも生体分子自身についての必要な情報をまとめて保存するとよい。例えば、分子名称、分子略号、構造コード

、機能コード、生物種、生成臓器、存在臓器等に関する情報を保存するのが望ましい。また、実験的に単離や存在の確認がされていない生体分子であっても、例えば他の生物種での実験から存在が推定される分子に仮に分子略号その他の情報を与えて定義してもよい。

【0033】

「生体分子情報データベース」には、各生体分子のアミノ酸配列や構造に関する情報を含めてもよいが、該情報を配列データベースや構造データベースに別途保存し、分子略号に基づいて必要に応じて取り出すのが望ましい。生体分子のうち低分子量のものについては、必要に応じて分子機能ネットワークの表示に化学構造を加えることができるように、正式な分子名称だけでなく、化学構造の表記に必要なデータを生体分子情報データベース又は別途のデータベースに保存しておくことが望ましい。

【0034】

2以上の生体分子が多量体又は集合体として活性を示す、或いは機能するなど、複数の生体分子をまとめて扱う方が便利な場合には、それらを1ヶの仮想的生体分子として定義し、分子略号を付して「生体分子情報データベース」に登録してもよい。その場合、その構成分子が既知の場合には、それぞれに分子略号を付して登録しておき、仮想的生体分子のレコードに構成分子の分子略号を記述するフィールドを設けるとよい。どのような生体分子で構成されているか不明な場合でも、集団として特定の機能をもつ仮想的生体分子を定義して、生体分子対の定義に用いることが可能である。

【0035】

また、生体分子が2以上のドメイン構造からなり、かつそれらが互いに異なる機能を持つなどの理由で各ドメインを独立に扱う方が有利と判断された場合には、各ドメインを独立の分子として扱ってもよい。例えば、元の生体分子と共に、各ドメインに分子略号を付して生体分子情報データベースに登録するとよい。元の生体分子のレコードに、分割したドメインの分子略号を記述するフィールドを設けることで、1ヶの生体分子が2以上の異なる機能をもつことが記述できる。遺伝子ではないゲノム配列上の特定の配列が、ある機能を持つ場合や特定の生体分

子によって認識される場合には、その配列の部分を独立の生体分子として扱い、分子略号を付して生体分子対の定義に用いることができる。

【0036】

生体分子対についての情報は「生体分子連鎖データベース」に保存する。生体分子対毎に、対をなす2つの生体分子の分子略号、関係コード、関係機能コード、信頼性コード、生体イベント、作用臓器、共役分子、その他の付加情報等を収録しておく。キイ分子とその標的生体分子の分子対については、生体イベント、両分子の量的又は質的な変動に対応した生体イベントのアップオアダウンの情報、病態イベント等の情報をできる限り入力しておくことが望ましい。キイ分子以外の生体分子対についても、該生体分子対が発現に直接関わる生体イベント又は病態イベントがある場合には、その生体イベントや病態イベントを入力しておくのが望ましい。キイ分子の量的又は質的な変動に対応した生体イベントのアップオアダウンの情報としては、例えば、キイ分子の増加に対して生体イベントが正常な範囲に比べて亢進するか低下するか等の単純化した情報で示してもよい。1ヶの酵素が2種以上の基質の反応を触媒してそれぞれ異なる反応生成物を生成する場合には、酵素と基質と反応生成物の関係を特定するための表現を付け加える。

【0037】

「生体分子情報データベース」と「生体分子連鎖データベース」は、その内容や構成が異なるため、本明細書では概念上独立のデータベースとして扱っているが、本発明の趣旨からも、双方を合わせた1ヶのデータベースに2種類のデータを含めても良いのは言うまでもない。また、「生体分子情報データベース」と「生体分子連鎖データベース」はそれぞれ2以上存在していてもよく、この場合、各データベースを適宜選択又は組み合わせて用いることができる。例えば、特定のフィールドで区別した、異なる生物種についてのデータを同じ「生体分子情報データベース」と「生体分子連鎖データベース」に保存してもいいし、ヒトとマウスについて別々のデータベースを作成して保存してもかまわない。

【0038】

「関係コード」としては、該生体分子対を構成する2分子がアゴニストと受容体である、酵素と基質である、というように単語で入力してもよいが、アゴニスト

と受容体の関係なら10, 酵素と基質の関係なら21, 酵素と生成物の関係なら22, といった具合に類型化して入力するのが望ましい。また、「関係機能コード」として、加水分解、リン酸化、脱リン酸化、活性化、不活化等の機能の別を保存しておくと便利であるが、これも類型化して入力しておくのが望ましい。

【0039】

酵素と基質のように互いの関係が明確になっている場合だけでなく、例えばtwo-hybrid実験法で蛋白-蛋白相互作用があるとされた2つの蛋白分子のように、両者の相互の役割が明確でない場合がある。このような生体分子対を含めてコネクト検索を行うためには、生体分子対をなす2分子の関係に方向性があるかないかを区別して扱うと便利である。各生体分子対に対して、どちらのケースに属するかを区別できるような関係コードを用いるのが望ましい。前者の場合は、作用方向確定として分子対の表現における2分子の入力順序のみが検索時に考慮されるのに対して、後者の場合は、検索時に作用方向不明として逆方向の関係も考慮される。このように生体分子対に方向性を持たせることによって、コネクト検索を行う際に、例えば分子機能ネットワークの上流側のみ又は下流側のみといった検索の向きを指定することも可能になる。

【0040】

直接結合する生体分子対の情報も、実験的に確実に証明されているものから、とりあえず生体分子対であると仮定されているものまで様々ある。また、実験法によってはfalse positiveのために間違っ生体分子対であるとされる場合も含まれる。そこで、各生体分子対の情報に信頼性のレベルや実験法を示す「信頼性コード」を付加しておくのが望ましい。信頼性コードは、単なる情報として閲覧するだけでなく、コネクト検索の際に、例えばある一定以上の信頼性を表わすコードを持つ生体分子対のみを対象とする等のフィルターをかける目的にも利用できる。また、分子機能ネットワークに信頼性コードに基づいたスコア付けを行って、結果の表示に反映させたり、表示する範囲の絞込みを行うなどの目的にも利用できる。

【0041】

生体分子が生成される臓器とは別に、それが蓄積され存在する臓器及び作用する

臓器の情報を保持しておく、分子機能ネットワークの生成時に、例えばある臓器で生成されて細胞外に出た分子が他の細胞の膜上にある標的生体分子に細胞外から作用する、といった現象を容易に表現できる。生体分子の生成臓器と存在臓器の情報は「生体分子情報データベース」に、作用臓器の情報は「生体分子連鎖データベース」に入力しておくのが望ましい。ここで生成臓器・存在臓器・作用臓器の記述は臓器に限定される必要はなく、組織、臓器又は組織内の部位、臓器又は組織内の特定の細胞、細胞内の部位などの情報を含んでいてもよい。作用臓器の情報は、コネクト検索を行う際に例えば同じ臓器内で生成する生体分子対のみを検索の対象とする等のフィルターとして利用することができる。

【 0 0 4 2 】

直接結合することを実証した実験や推定の方法、生体イベントの種類、キイ分子の量的変動に対応した生体イベントのアップオアダウン、細胞内部位や組織、臓器、臓器内部位の表現は単純化してあれば何でもよいが、類型化して短い英数字記号等に変換しておくのが望ましい。同義語辞書で定義しておけば、同時に同義語の処理もでき、入力時のミスが最小化できるからである。

以下に「生体分子連鎖データベース」から「分子機能ネットワーク」を生成する「コネクト検索」の概念を示す。本発明の「コネクト検索」には、この概念を実現できるものであれば、いかなる方法を用いてもよい。例えば、Sedgewickによる「アルゴリズムC（近代科学社、1996）」第29章に記述されている「深さ優先探索」のアルゴリズムなどが利用できる。

【 0 0 4 3 】

分子略号a~z で表される生体分子からなる各生体分子対を(n,m) のように表現するとすると、生体分子連鎖データベースは次のような生体分子対の集合として表される。

(a, c) (a, g) (b, f) (b, k) (c, j) (c, r) (d, v) (d, y) (e, k) (e, s) (g, u) (j, p) (k, t) (k, y) (p, q) (p, y) (x, z)

コネクト検索で、例えばcとeを含む分子機能ネットワークを生成するよう指定したとすると、対のうち的一方の分子を共通にもつ生体分子対 (c, j) (j, p) (p, y) (y, k) (k, e) を次々に探し、分子機能ネットワークとして、分子c, j,

p, y, k, e のつながりである「c - j - p - y - k - e」を得る。

【0044】

得られた「分子機能ネットワーク」に基づいて、生体イベントの推定を以下のように行う事ができる。生体分子eがキイ分子であって生体イベントEの情報をもつ場合、生体分子c, j, p, y, k は生体イベントE の発現に直接又は間接に関係すると推定できる。さらに、例えば分子eが減少するとEの発現が亢進するといった生体イベントのアップオアダウンの情報がある場合には、(c, j) (j, p) (p, y) (y, k) (k, e)のそれぞれの関係を考慮して、c, j, p, y, k のうちの任意の分子の量的又は質的変動に対する生体イベントEの発現への影響を推定することができる。

【0045】

コネクト検索は、生体分子を3以上の複数個指定して行うこともできる。例えば a, b, c の3つの生体分子が指定された場合に、aとb, aとc, bとcの3組についてそれぞれコネクト検索を行い、

a - d - e - f - b,

a - d - e - g - c,

b - f - e - g - c

の3組の分子機能ネットワークを得たとする。これらの共通な部分（AND演算）や和集合（OR演算）を求めたり、2組以上に共通する部分を求めたりすることによって、a, b, c の3つの生体分子を含む分子機能ネットワークを得ることができる。

【0046】

a, b, c の3つの生体分子が全て生体イベントEの情報を持つ場合、図2に示すようにOR演算により和集合によって求められる分子機能ネットワークに出現する生体分子群a, b, c, d, e, f, gは、生体イベントEの発現に直接又は間接に関係すると推定できる。さらに、AND演算により共通な部分として求められる分子機能ネットワークに出現する生体分子eは、生体イベントEの発現に対して重要な役割を果たしている可能性が高いものと推定できる。

【0047】

コネクト検索は、サブネット、病態イベント、医薬分子、遺伝子、疾患に関する情報などのデータ項目間の関連付け情報を用いて、上記に示した生体分子対の情報を用いる場合と同様に行うことができる。この場合、関連付け情報中で関連付けされた2つのデータ項目を、生体分子対中の2つの生体分子と同様に扱って上記の手順を行えばよい。これにより、生体分子と生体イベントのみならず、サブネット、病態イベント、医薬分子、遺伝子、疾患に関する情報などを指定して分子機能ネットワークを生成することが可能になる。

【0048】

さらに、ある生体分子からキイ分子までの分子機能ネットワーク上にあるNヶの生体分子が、生体イベントの発現量 Q_E に与える影響を、例えば次のような式で予測することもできる。ここで、 S_i はi番目の生体分子の状態の質的な評価値、 R_i はi番目の生体分子の量を表わす値、 V_i はi番目の生体分子が存在する環境の評価値を表わし、 f は $3 \times N$ ヶの入力値をもつ多価関数である。

$$Q_E = f(S_1, R_1, V_1, \dots, S_N, R_N, V_N)$$

【0049】

1ヶの分子機能ネットワークが関係する生体イベントは1種類に限らないし、また、1種類の生体イベントに関わる分子機能ネットワークが数ヶあることが予想されるが、生体イベントの側から関連のある分子機能ネットワークを絞り込むことは可能である。例えば、1以上の生体分子を指定して「分子機能ネットワーク」を生成した場合に、著しく多数の生体分子が含まれる「分子機能ネットワーク」が生成された場合には、生体イベントの情報を加えて「分子機能ネットワーク」の範囲を絞り込むことが可能である。当然のことながら、何らかのメディエーター分子、又は該分子と標的生体分子の関係を含むことを条件として、「分子機能ネットワーク」を生成させることも可能である。

【0050】

また、「生体分子連鎖データベース」のデータを適宜分割し、又はフィルターをかけ、又は部分集合を抽出し、又は階層化することによって、必要な範囲の分子機能ネットワークを生成することができる。分割やフィルターや部分集合の抽出は、本発明のデータベース特有のデータ項目に対する検索、キーワードを用いた

一般的な文字列検索、アミノ酸配列又は核酸配列に対する相同性検索、化学構造式に対する部分構造検索、生体分子対の方向性を指定した検索、本発明のデータベース特有のデータ項目を一定の基準に沿ってスコア付けに利用した検索等の検索方法により行なうことができる。これらの検索を予め「生体分子連鎖データベース」又は「生体分子情報データベース」に対して行なうことにより、限定をかけた分子機能ネットワークや特徴づけした分子機能ネットワークを生成することができる。

【0051】

例えば、生成臓器や作用臓器の情報を用いて、肝臓で生成する生体分子とか皮膚で起こる生体イベントといった観点から絞り込んだ部分データベースを作成してコネクト検索することで、限定した範囲の「分子機能ネットワーク」を生成することができる。また、コネクト検索で生成させた分子機能ネットワークに対して、そこに含まれる生体分子又は生体分子対について上記の検索を行なって分割やフィルターや部分集合の抽出をすることによって、所望の特徴をもつ分子機能ネットワークや所望の範囲の分子機能ネットワークを生成することができる。さらに、コネクト検索を行う際に例えば生体分子対の方向性を指定してネットワークの向きを絞った検索を行ったり、コネクト検索した結果の分子機能ネットワークを信頼性コードに基づくスコア付けを行って絞り込むこともできる。こうした限定や特徴づけは検索を容易にするだけでなく、特定の生体分子群や生体分子対を分子機能ネットワーク上で強調して表示して、分子機能ネットワークの理解を助ける目的にも有効である。

【0052】

「生体分子連鎖データベース」の分割やフィルターや部分集合の抽出を、ネットワーク上の連なりに基づいて適切に行ない、その包含関係を示す情報を保存して利用することにより、「分子機能ネットワーク」の階層化が可能になる。未解明の分子や分子間のつながりが一部にあっても、それらをまとめて1ヶの仮想的な生体分子として他分子との対を定義し、仮の分子機能ネットワークを生成することも可能である。含まれる分子数が多いために過度に複雑なネットワークが生成される場合には、ネットワーク上で連結している2以上の生体分子群を仮想的な

1ヶの生体分子と定義して、ネットワークをシンプルに表現することも可能である。

【0053】

このような階層化を利用することにより、コネクト検索の高速化を可能にし、ネットワーク表示の詳細さを調節可能にして、過度な複雑さを適切に回避できる。本明細書中では、このようにネットワーク上で連結している2以上の生体分子対からなる部分的なネットワークを「サブネット」とよぶ。

【0054】

サブネットとしてはいかなる部分ネットワークを指定してもよいが、好ましくは代謝系におけるTCA回路やペントースリン酸回路の例のように研究者によく知られたカスケード、パスウェイ、回路等をサブネットとするのが便利である。また、あるサブネットが別のサブネットに含まれていてもよく、例えば代謝系自体を複数のサブネットを含む上位のサブネットと見なすこともできる。

【0055】

各サブネットを1ヶの仮想生体分子として扱う方法もあるが、サブネットを構成する生体分子対とサブネットの階層に関する情報を「生体分子連鎖データベース」に保存するのが便利である。また、「生体分子連鎖データベース」にサブネットを表わすための上位のデータ階層を設けて、そこに該サブネットについての情報を保存してもよい。生体分子対のサブネットによる階層化は2層に限らず、複数のサブネットの集合をより上位のサブネットとして保存してもよい。分子機能ネットワーク生成時に個々の分子対のデータと上位階層のサブネットデータの間での相互参照を容易にするために、個々の分子対データ・サブネットデータそれぞれに互いの関連を示す情報を格納しておくのが望ましい。1ヶの生体分子対が複数のサブネットに関連付けられてもよいことは言うまでもない。

【0056】

階層化した「生体分子連鎖データベース」のサブネットのデータには、下位階層の生体分子対との関連のみならず、サブネット間の関連の情報を含めるのが望ましい。例えば、代謝系において解糖系とTCA回路は連続して動作するサブネットであり、これらのサブネット間の関係を上位階層における対の関係として保存す

ることができる。この場合、サブネットの対の情報に加えて、サブネット間の接点となる生体分子の情報を加えておくことが望ましい。

【0057】

さらに、ネットワークの階層化に加えて、生体分子自身も階層化でき、「生体分子情報データベース」にその情報を保存して利用できるのが本発明の特徴である。ネットワークの高速な検索と便利で多様な表示のためには、生体分子情報と生体分子対情報の双方を階層化しておくのが望ましい。生体分子の階層化の対象としては、以下のような例を挙げることができる。生体分子のなかには、複数の異なる分子が特異的に集合してある機能を示すものがあり、分子の集合状態の違いにより機能の発現状態や種類が制御されている場合も多い。また、免疫細胞などで見られるように、細胞表面に発現している複数の分子の組み合わせにより、生体イベントとの関連や細胞の機能が規定されている場合もある。このような場合に、上記のように分子の集合状態を1ヶの仮想的生体分子として扱う方法もあるが、別の方法として、「生体分子情報データベース」に分子の集合状態を表わすための上位のデータ階層を設けて、そこに該集合状態についての情報を格納してもよい。分子機能ネットワーク生成時に生体分子データと上位階層のデータの間の相互参照を容易にするために、生体分子データと上位階層データのそれぞれに、互いの関連を表わす情報を格納しておくのが望ましい。1ヶの生体分子が複数の上位階層データに関連付けされてもよいことは言うまでもない。

【0058】

生体分子の集合状態を階層化により扱う別の方法として、生体分子情報データベースの分子の下位の階層に、集合に加わっている各分子を登録する欄を設けてもよい。この場合、集合状態にある生体分子を一つの仮想的生体分子として登録し、上記で設定した下位の階層の欄に各分子の分子略号を登録するのが好ましい。

【0059】

生体分子のなかには、酵素反応などによる修飾を受けることにより、その機能が変化するものがある。このような修飾の例としては、リン酸化、メチル化、アセチル化、ユビキチン化やそれらの逆反応などがある。修飾状態が異なる生体分子は、別々の分子略号を与えることにより区別してもよいが、生体分子情報データ

ベースに修飾の情報を示すデータ階層を設けるとより便利である。例えば、生体分子情報データベースに「修飾状態」を表わす欄を設けて、そこに「リン酸化」などの記述をすることにより、異なる修飾状態にある生体分子であっても、同一の分子略号で扱いつつ、修飾状態を区別することが可能になる。このような修飾状態による階層化を行なった場合には、該分子を分子略号と「修飾状態」欄のデータの組合せにより区別して扱えばよい。

【 0 0 6 0 】

上記の階層化は、生体分子の修飾状態以外にも拡張可能である。例えば、生体分子の構造変化や複数の修飾の組合せ等によって、該分子の他の生体分子への結合性や反応性が変化するような場合には、その状態変化を例えば活性型・不活性型として区別して分子データの下位の階層に保存するとよい。

【 0 0 6 1 】

生体分子によっては、修飾状態や活性型・不活性型などの状態変化に依存して他の生体分子と結合したりしなかったりするものがある。このような場合には、上記の階層化した分子データに基づいて生体分子を区別して扱い、上記の状態毎に生体分子対を生体分子連鎖データベースに記録してもよい。また、別の方法として、生体分子対をとりあえず分子略号のみに基づいて生体分子連鎖データベースに記録しておき、該生体分子対への付加情報として該2分子が結合・反応する条件となる各分子の状態を記録してもよい。

【 0 0 6 2 】

生体イベント・病態イベントのなかには、特定の生体分子対に関連付けすることのできないものも多い。例えば、生体イベント・病態イベントとあるサブネットの形成との間の関係は分かっているが、該イベントに直結する生体分子対が未知の場合がある。このような場合には、上記の生体分子対データの階層化を利用して、生体イベント・病態イベントを生体分子対の上位階層であるサブネットのデータに対して関連付けることで、該イベントと生体分子ネットワークの関係を記述することが可能となる。

【 0 0 6 3 】

また、特定の分子の集合状態や細胞表面への特定の分子の発現状態がある生体イ

イベント・病態イベントの発現に関連している場合には、上記の分子集合状態や分子発現状態の階層化を利用して、生体イベント・病態イベントを分子集合状態や分子発現状態の階層のデータに関連付けておくことで、該イベントと生体分子ネットワークの関係を記述することが可能となる。

【0064】

さらに、生体イベント・病態イベントのなかには、特定の生体分子対やサブネットのいずれにも関連付けすることができないものもある。このような例の一つとして、炎症性サイトカインの遊離・白血球の組織への浸潤・毛細血管の透過性の向上などの様々な生体イベントが組み合わさって起こる「炎症」という病態イベントを挙げることができる。このようなイベントを扱うためには、生体イベント・病態イベントを階層化し、下位の階層には生体分子対やサブネットに関連付けられるイベントを記述し、上位階層には下位階層のイベントに関連して起こるイベントを記述するとよい。この階層化では、2層以上の階層構造を用いてもよいことは言うまでもない。各階層間でのイベントの相互参照を容易にするために、各階層のイベントのデータに上下の階層のデータへの関連を示す情報を格納しておくのが望ましい。このような生体イベント・病態イベントのデータの階層化により、特定の生体分子対やサブネットに直接関連付けられないイベントについても、分子機能ネットワークとの関係を記述することが可能となる。

【0065】

生体イベント・病態イベントの階層化の別の態様として、生体イベント・病態イベントをキイ分子のアップオアダウン、生体反応、症状・症候、臨床検査値の情報、疾患・合併症などからなる複数のグループかつ／又は階層に分けてもよい。例えば、ある疾患において、上記のグループ又は階層に分けられた生体イベント・病態イベントの間に関連が見られる場合に、該イベントの間に関連付けをしておくといふ。この方法によると、キイ分子のアップオアダウンの情報を經由して、疾患と分子機能ネットワークとの関係を容易に記述することができる。

【0066】

以上例示したように「生体分子情報データベース」及び「生体分子連鎖データベース」のデータを階層化して保存することにより、多様な用途に対応した分子機

能ネットワークの生成を効率的に行なえるようになる。

同一の生体イベント・病態イベントであっても、例えば正常時と疾患時や、疾患の状態の違いなどによって、関連付けられるべき生体分子対やサブネットが変化する場合もある。このような場合には、該イベントからそれぞれの生体分子対やサブネットに対して関連付けをしておくことで、同一の生体イベント・病態イベントが関わる2以上の分子機能ネットワークを探索することが可能となる。

【0067】

解糖系に存在するある生体分子（分子A）とあるキナーゼカスケードに存在するある蛋白（分子B）との関連を調べようとする場合に、階層化していないデータを用いる方法では膨大な数の分子対を対象にコネクト検索を行なう必要があり、分子Aと分子Bの間の経路が長い場合には事実上検索が不可能となる。一方、階層化されたデータを用いると、「解糖系」というサブネットと「あるキナーゼカスケード」というサブネットとの間の関連をサブネットという上位階層においてコネクト検索し、上位階層での経路が見つかった場合には、必要に応じて経路上にある各サブネットの下位階層でコネクト検索を行なうことができる。このように経路探索問題を異なる階層の問題に分割することにより、階層化を用いない場合には不可能であったような分子機能ネットワークの生成も可能となる。この階層化データを利用したコネクト検索を行う際にも、階層化を行わないデータを使ったコネクト検索と同様に、サブネット間の関係の方向性又は本発明のデータベース特有のデータ項目をそのまま又は一定の基準に沿ってスコア化して検索対象データを絞り込むこともできる。

【0068】

また、上記の階層化データを利用したコネクト検索において、特定のサブネットが頻繁に参照されるような場合には、該サブネット内について予めコネクト検索を行なっておき、該サブネット内の分子機能ネットワークの情報を保存しておくといふ。この処理により、全体の分子機能ネットワークの生成をより効率的に行なえるようになる。

【0069】

さらにまた、例えば、「炎症」という病態イベントに関わる分子機能ネットワー

クを生成する場合に、「炎症」という上位階層のイベントに関連する下位階層のイベントを探し、該下位階層イベントが関連する生体分子対又はサブネットから出発してコネクト検索することにより、より広範囲な分子機能ネットワークの生成が可能となる。

【0070】

上記のように、本発明によると生体分子間の直接結合の情報の情報から、任意の分子に関連した分子機能ネットワークが生成でき、直接又は間接に関わる生体イベント及び病態イベントが容易に推定できる。また、本発明は逆に、疾患で特徴的に現れる生体イベントや病態イベント、生体分子の量の変化などの知見から、疾患に関連する可能性の高い分子機能ネットワークを選別し、疾患の分子的メカニズムを推定する目的に利用できる。さらに、本発明によると、特定の疾患や症状の治療にネットワークのどの過程を阻害するのが有効か、ネットワーク中のどの分子が創薬ターゲット（医薬開発で標的とする蛋白質その他の生体分子）として有望か、その創薬ターゲットから予想される副作用は何か、それを回避するにはどのようなアッセイ系で医薬開発候補を選別するかと言った創薬戦略の構築が可能になる。

【0071】

医薬分子は一般に、体内で蛋白質をはじめとする生体高分子に結合してその機能を制御することにより薬理活性を発現する。それらの分子の作用は生体分子の作用よりも詳しく研究され、対象疾患の分子メカニズムの解明に役立ってきた。そこで、製造承認を受けて医療に用いられている医薬分子や薬理学研究等に用いられる薬物分子とその標的生体分子の対の関係を、上記の生体分子及び生体分子間の情報に加えることにより、本発明の方法の有用性を高めることができることに注目した。標的生体分子は殆どの場合、蛋白質又は糖などで修飾された蛋白質である。標的生体分子を含む分子機能ネットワークから、副作用となりそうな生体イベントを推定することができるし、併用する薬剤の関わる分子機能ネットワークとの交差から、薬剤間相互作用を推定することも可能になる。その結果、副作用リスクや薬剤間相互作用のリスクを考慮した薬剤の選択や用量の設定が可能になる。

【0072】

以下に本発明で医薬分子と標的生体分子の対の関係を加えた場合の方法の例を示す。医薬分子の正式名称に対して、分子略号を定義し、該分子そのものについてのすべての情報を収録する「医薬分子情報データベース」を作成する。ここには医薬分子の名称、分子略号、適用疾患、用量、標的生体分子その他の情報を保存する。生体分子情報データベースの場合と同様に、医薬分子の化学構造、アミノ酸配列（ペプチド又は蛋白質の場合）、立体構造等の情報を「医薬分子情報データベース」に含めてもよいが、別途のデータベースに保存することが望ましい。医薬分子と生体分子、蛋白質と低分子などを区別する目的には、構造コード等で区別してもよいし、最初の1文字で区別できるような分子略号の付け方を採用してもよい。さらに、医薬品の添付文書やその他の文献から、顕著な副作用、他の薬剤との相互作用、代謝酵素等の情報も入力しておく、分子機能ネットワークに基づいて、遺伝子多型との関連で薬剤の適正な選択等を行う目的に役立つ。

【0073】

さらに、医薬分子と標的蛋白質の対とその関係に関する情報を含むデータベースである「医薬分子連鎖データベース」を作成する。医薬分子の分子略号、標的生体分子の分子略号、関係コード、薬理作用、適用疾患その他の情報を保存する。標的生体分子の分子略号については、生体分子情報データベース中で定義したものをを用いる必要がある。関係コードなど生体分子連鎖データベースと共通するデータ項目については、生体分子連鎖データベースに準じた表記方法を用いるとよい。

【0074】

「医薬分子情報データベース」と「医薬分子連鎖データベース」を作成して医薬分子及び医薬分子対の情報を取り込むことにより、図3に示すように本発明の方法を拡張することができる。ここでのコネクト検索による分子機能ネットワークの生成や生体イベントの推定などは、上記の生体分子連鎖データベースと生体分子情報データベースだけを用いた場合と同様の方法で行うことができ、該ネットワーク上の分子を標的とする既存医薬分子の情報が同時に得られる。また、生体分子連鎖データベースと生体分子情報データベースだけを用いて作成した分子機

能ネットワークから、指定した医薬分子の関わる分子機能ネットワークを抽出する目的にも利用できる。

【0075】

一方、ヒトゲノム配列の解析をはじめとして、多様な角度から遺伝情報の解明が急速に進みつつある。ゲノムワイドにcDNAが単離され、orf (open reading frame) や遺伝子配列の解明が進み、各遺伝子のゲノム上への位置づけが進んでいる。そこで、本発明の別の態様として、生体分子のうちの蛋白質の分子略号と、該蛋白質をコードする遺伝子の名称、略称、IDその他の情報を関係付けた生体分子-遺伝子データベースを作成しておくことにより、以下のように本発明の方法を拡張することができる。すなわち、遺伝子と生体分子を対応づけることにより、疾患のマーカーとなる遺伝子や蛋白の意味付け、疾患と遺伝子多型の関わり等の知見を分子機能ネットワーク上の分子及び生体イベントとの関連において理解することが可能になる。生体分子-遺伝子データベースには、生物種、ゲノム上の位置、遺伝子配列、機能のほか、遺伝子多型のアミノ酸変異と略称、機能との関連等の情報を含めておくのが望ましく、必要に応じて2以上のデータベースとするのもよい。

【0076】

ゲノム配列上に位置づけられた遺伝子名や遺伝子の並びから、特定のキイ分子が核内受容体に作用することによって転写される蛋白が明確になり、生体分子間の相互の制御の関係を分子機能ネットワークに反映することが可能になる。また、臓器によって発現する遺伝子や蛋白が異なることが知られているが、本発明の方法によれば、そうした発現情報を「生体分子情報データベース」に取り込むことにより、臓器毎に異なる「分子機能ネットワーク」を生成することができ、例えば、異なる臓器では核内受容体を標的とする医薬分子の作用が違ったり逆転したりする現象の説明が可能になる。また、医薬分子を投与した場合に蛋白発現が変化する事実がわかっているが、本発明の方法により個々の発現蛋白量の増減を標的的生体分子に関わる分子機能ネットワーク上で解釈すると、遺伝子多型を考慮した薬剤選択に有用である。

【0077】

上記の遺伝子転写や蛋白発現の情報の保存においても、階層化の概念を利用することにより、より効率的かつ広範囲な分子機能ネットワークの生成が行なえるようになる。例えば、特定の核内受容体により転写・発現される複数の遺伝子・蛋白に対して、「生体分子情報データベース」に遺伝子群の転写・蛋白群の発現を表す上位の階層を設けて、そこに該遺伝子群・該蛋白群のデータを保存しておくといふ。該遺伝子群の転写や該蛋白群の発現に関連する生体イベント・病態イベントが存在する場合には、「生体分子連鎖データベース」に該遺伝子群・該蛋白群の上位階層データと該イベントとの関連を記載することにより、個々の遺伝子・分子と該イベントとの間の関連付けでは表せないような分子機能ネットワークの生成が可能となる。

【0078】

上記の遺伝子転写や蛋白発現の情報の階層化した保存法において、該遺伝子群の個々の遺伝子又は該蛋白群の個々の蛋白の転写又は発現の量的な情報が分かっている場合には、それらの情報も数値パラメタとして「生体分子情報データベース」に保存しておくといふ。これらの数値パラメタを利用することにより、個々の遺伝子の転写量又は個々の蛋白の発現量の違いに応じて、関連する生体イベント・病態イベントが変化する場合は記述することが可能となる。

【0079】

また、ゲノムや遺伝子についての個体の多様性が明らかにされつつあるが、そうした情報を本発明の方法とリンクすることにより、個体差への理解が進み、個体差に基づく治療が可能になる。特定の生体分子（蛋白質）の機能が損なわれるような遺伝子多型について、分子機能ネットワーク上で解釈することにより生体イベントへの影響が推定できる。1ヶの遺伝子の欠損や異常が引き起こす遺伝病の症状及び生体イベントの異常の情報を、本発明の方法にリンクさせることも、その理解に有用である。

【0080】

いくつかの代表的疾患で、各疾患の患者に頻度高く見出される数ヶの遺伝子、疾患背景遺伝子の存在が報告されている。特定の疾患にかかりやすい遺伝的体質が実際にあると仮定した場合、例えば、血圧の調節に関わる分子機能ネットワーク

は2以上あり、そのうちのどのネットワークのどの分子の異常かによって、高血圧の背景となる遺伝子は相当数あっても不思議ではない。こうしたpolygenicな遺伝子の問題を解釈するには、本発明の方法が不可欠である。

【0081】

また、近年、マウス、ラット等の動物のゲノムや遺伝子の解析も急速に進められており、ヒトのゲノムや遺伝子との対応も付けられるようになってきた。これらの動物とヒトとで生理機能の調節に関わる蛋白がかなり似ていることが期待されるが、かなりの違いもあることが医薬開発の障害となってきた。これらの動物とヒトの間で蛋白や蛋白機能が大きく異なる場合があることがわかってきており、本発明の方法とのリンクによって、ヒトの分子機能ネットワークとの違いを明らかにすることは医薬開発に有用である。さらに、一般にヒト用に開発された医薬を転用することが多い動物薬について、適切な使用を目指す上でも有用である。

医薬開発においてはヒトの疾患と病態が類似した病態動物がある場合、その動物での薬理活性を指標に開発を進めることが多い。そうした病態動物の遺伝子も進められており、本発明の方法によりヒトの遺伝情報と対応づけることによって、ヒトの該疾患のメカニズムの解明に役立つ。

【0082】

さらに、遺伝子機能の解明の目的で、特定の遺伝子を潰したノックアウト動物や機能の弱い遺伝子に変えたり遺伝子を過剰発現させたりしたトランスジェニック動物を作成することが多くなった。これらは致死的で生まれて来ない場合や生理機能や行動に何の影響も見られない場合も多いが、生まれてきた動物に何らかの異常が観察される場合でも、これらの動物実験の結果の解析は非常に難しいといわれている。このような実験において、本発明の方法を用いて、該遺伝子操作の影響を予測してから、機能解析すると便利である。

【0083】

遺伝子関連の情報は、解析の進展とともに、配列IDの面からの統合の試みも進みつつあり、さらにゲノム配列上に遺伝子を位置づける試みも進められている。上記「生体分子連鎖データベース」との連携を考慮した独自の遺伝情報データベースを構築して上記の目的に利用することも考えられるが、これらの情報が膨大

でかつ公開される方向にあることを考えると、将来はそうした公開の情報を本発明の方法に臨機に取り込むことによって、上記の方法が実施できる可能性が高い（図4）。

【0084】

本発明の方法で用いる生体分子連鎖データベースは、必ずしも同一の場所にまとめて管理・保存されている必要はなく、分子略号を統一することにより、異なる場所に管理・保存されている1以上の生体分子連鎖データベースを適宜選択して通信等の手段により接続して用いることもできる。生体分子連鎖データベースのみならず、本発明の方法で用いる生体分子情報データベース、医薬分子連鎖データベース、医薬分子情報データベース、遺伝情報データベースなどについても、同様の扱いが可能なのは言うまでもない。

【0085】

本発明の実施のさらに別の態様として、生体イベントの発現に直接関わる生体分子と該生体イベントの情報を含むデータベース（生体イベントー生体分子データベース）を作成し、生体イベントの情報を必ずしも含まない分子ネットワークのデータベースと併用する方法も提供される。さらに別の実施の様態として、生体イベントの情報を必ずしも含まない分子ネットワークのデータベースから、任意の分子に関わる部分的な分子ネットワークを抽出し、該ネットワークを構成する分子に基づいて、上記の生体イベントー生体分子データベースを検索する方法も提供される。

【0086】

本発明の実施のさらに別の態様として、「生体分子情報データベース」・「生体連鎖データベース」・「医薬分子情報データベース」・「医薬分子連鎖データベース」・「生体分子ー遺伝子データベース」などの各データ項目に対して、キーワード・数値パラメタ・分子構造・アミノ酸配列・塩基配列などに基づく検索を行ない、該検索結果に基づいて分子機能ネットワークを生成する方法が提供される。以下に、検索に基づく分子機能ネットワーク生成の例を挙げるが、本発明の範囲がこれらの例に限定されるものでないことは言うまでもない。

【0087】

各データベースには、分子名称・分子略号・生物種・生成臓器・存在臓器等の種々の情報がテキストとして保存されている。これらのテキストに対して文字列の全一致・部分一致などに基づく検索を行なうことで、生体分子・生体分子対・生体イベント・病態イベント・医薬分子・医薬分子-生体分子対・遺伝子-蛋白の対応データなどを絞り込むことができる。これらの絞り込まれた情報に基づいて、コネクト検索の起点・終点を定めたり、コネクト検索で対象とする分子対の範囲を狭めたりすることができ、利用目的に応じた分子機能ネットワークの生成が可能となる。

【0088】

「医薬分子情報データベース」に医薬分子の化学構造・立体構造が保存されている場合には、これらに対して全体構造の一致・部分構造の一致・構造類似性などに基づく検索を行なうことで、医薬分子を絞り込むことができる。絞り込まれた医薬分子に基づいて、該医薬分子に関連する分子機能ネットワークを生成したり、該医薬分子に関連する生体イベント・病態イベントを検索したりすることが可能となる。

【0089】

「生体分子情報データベース」に遺伝子転写・蛋白発現等の数値パラメタが保存されている場合には、これらの数値パラメタに基づく検索を行なうことで、遺伝子の転写量・蛋白発現量等に応じた分子機能ネットワークの生成が可能となる。

「生体分子情報データベース」又は関連データベースに蛋白のアミノ酸配列が保存されている場合には、これらのアミノ酸配列に対して配列相同性や部分配列パターンの一致などに基づく検索を行なうことで、生体分子を絞り込むことができ、該生体分子に基づいて分子機能ネットワークを生成することができる。この方法は、機能未知の蛋白質やその部分配列情報について、該蛋白質が関係する可能性が高い分子機能ネットワークを推定し、さらに該蛋白質の機能を推定する目的に有効である。

【0090】

「生体分子情報データベース」、「生体分子-遺伝子データベース」又は関連データベースに蛋白に対応する遺伝子の塩基配列が保存されている場合には、これ

らの塩基配列に対して配列相同性や部分配列パターンの一致などに基づく検索を行なうことで、生体分子を絞り込むことができ、該生体分子に基づいて分子機能ネットワークを生成することができる。この方法は、機能未知の遺伝子やその部分配列情報について、該遺伝子から翻訳される蛋白が関係する可能性が高い分子機能ネットワークを推定し、さらに該蛋白の機能を推定する目的に有効である。

【0091】

本発明のさらに別の態様として、ある疾患（以下、「注目疾患」という）に関連する病態イベント、生体分子、その他の情報をグループ化かつ／又は階層化して保存した「病態連鎖データベース」を利用する分子機能ネットワークの生成・検索方法も提供される。以下、病態連鎖データベースの作成及び利用方法の例を示すが、本発明の範囲が以下の特定のグループ化・階層化の方法に限定されないの言うまでもない。

【0092】

病態連鎖データベースには、注目疾患に関する病態イベント、生体分子、その他の情報を以下の例のようにグループ化かつ／又は階層化して保存する。さらに、ある注目疾患において、いずれかのグループに属する2以上のデータ項目の間に関連が認められる場合には、該データ項目間の関連付けの情報を保存する。

1. キイ分子
2. 生体反応
3. 症状・症候
4. 臨床検査値
5. 疾患・合併症

病態連鎖データベースの「キイ分子」のグループには、疾患に関連して量的かつ／又は質的に変動が見られる生体分子の情報を保存する。このキイ分子の情報を利用することにより、病態イベントから分子機能ネットワークを検索したり、分子機能ネットワーク中の分子と病態イベントの間の関連を検索したりすることが可能となる。

【0093】

キイ分子の量的又は質的な変動は、以下に示す例のように表現することができる

。ある疾患において分子Aが量的に増加の方向に変動する場合をA(+)と表現する。
 。この例としては、酵素活性や受容体活性の亢進などが挙げられる。その反対に、分子Aが量的に減少の方向に変動する場合をA(-)と表現する。この場合には量的にはほとんど0に近い場合も含んでもよく、それに伴う分子のloss of functionをも含んでもよい。例えば、DNAのプロモーター領域のCpGアイランドのメチル化異常により遺伝子が発現せず、そのために本来転写・翻訳され、機能すべき蛋白質が不活化しているという現象もこの表現によって示すことができる。また、この表現方法を用いると、特定の遺伝子をロックアウトし、その遺伝子がコードする蛋白質が欠損した際の分子機能ネットワークの変化を推定することが容易となる。ある疾患において、ある分子が質的に変化する場合をA(m)と表現する。これはその蛋白質をコードする遺伝子の変異などによることが多い。

【0094】

病態連鎖データベース中の「生体反応」のグループには、直接的または間接的に疾患に帰着すると考えられる、生体にとっての異常または過剰（あるいは過少）な状況にある生体反応、生体内現象の情報を保存する。生体反応のグループに記述されるデータ項目は、分子機能ネットワーク上の生体イベント、生体分子対、サブネット等と関連付けしてもよい。ここで生体反応とは、主に分子レベル（蛋白質、酵素活性、遺伝子を含む）、細胞レベル、臓器レベルにおける現象を指し、さらに情報を階層化して記述してもよい。また、それが起こる場の情報や条件等により、生体反応を階層化することもできる。

【0095】

病態連鎖データベース中の「症状・症候」のグループには、注目疾患を特徴づける、あるいは注目疾患の診断の根拠となる病態イベントを保存する。医薬分子による副作用に対応する病態イベントも、このグループに記述するのが好ましい。また、このグループに存在する病態イベントのうちで、ある医薬分子の作用機序や適応症に対応又は関連するものについては、「医薬分子情報データベース」中で該医薬分子のデータとして記述しておくといよい。これにより、作用の標的となる生体分子が分かっていない医薬分子についても、分子機能ネットワークを検索することが可能になる。

【0096】

病態連鎖データベース中の「臨床検査値」のグループには、実際に臨床で用いられているか、用いられる可能性のある検査項目を保存する。臨床検査で測定される対象となる生体分子が生体分子情報データベース中に存在するときは、その生体分子の分子アノテーションとして該分子が臨床検査の測定対象であるという情報を付加してもよい。臨床検査値のグループ中の生体分子について、関与する分子機能ネットワークを検索することができ、分子機能ネットワーク中に存在する生体分子が臨床検査の測定対象となっている場合に、この情報に基づいて該生体分子を強調表示するなどの目的にも利用できる。

【0097】

病態連鎖データベース中の「疾患・合併症」のグループには、注目疾患の成因、ステージ、罹患部位などに基づく名称や分類名、及び注目疾患に関わる合併症などの情報を記述する。「疾患・合併症」のデータ項目は、例えば注目疾患、ステージ、罹患部位などに基づいて階層化して保存してもよい。病態連鎖データベース中の用語は、例えばMedDRA/J (Medical Dictionary for Regulatory Activities Terminology; 日本公定書協会) などの標準辞書に従って記述するのが好ましい。また、辞書に含まれている用語のほかに、類義語やシノニム等も付加しておくもよい。

【0098】

病態連鎖データベースと、本発明の他のデータベースを併せて用いることにより、以下のような分子機能ネットワークの検索が可能となる。上述のとおり、ある医薬分子に対して標的生体分子が記述されている場合には、該医薬分子から該標的生体分子を経由してコネクト検索を行なうことにより、該医薬分子が関わる分子機能ネットワークを検索することができる。病態連鎖データベースを用いると、標的生体分子が不明な医薬分子であっても、以下のようにして該医薬分子が関わる分子機能ネットワークを検索することが可能となる。

【0099】

標的生体分子が不明な医薬分子について、該医薬分子の適応疾患、適応症、作用機序等の情報に基づいて、病態連鎖データベース中で一致する項目を検索する。

次に、病態連鎖データベース中の関連付けの情報を用いて病態連鎖データベースのデータ項目間のコネクト検索を行ない、分子機能ネットワークにつながり得るキイ分子の情報を取得する。該キイ分子から、生体分子連鎖データベースを用いたコネクト検索を行なうことで、該医薬分子が関わる分子機能ネットワークを検索することができる。また、ある医薬分子の副作用情報に基づいて病態連鎖データベースを検索し、症状・症候や疾患・合併症のグループ中に一致する項目が存在すれば、その項目から上記と同様にコネクト検索を行なうことにより、該医薬分子の副作用に関連する分子機能ネットワークを検索することができる。

【0100】

本発明によって提供されるデータベース及びコンピュータシステムは、利用者が容易にデータの追加、編集、削除等を行うことができる入出力装置としても応用できる。利用者が編集したデータは、保存可能であり、必要に応じて予め提供されたデータベースとマージすることができる。さらに、マージしたデータは検索、閲覧等に利用することもできる。例えば、利用者は実験等によって実証または予測された生体分子、生体分子対、生体イベント、病態イベント、医薬分子情報、遺伝子、その他の情報に基づいてデータの追加、編集、削除を行うことにより、コネクト検索により新規な分子機能ネットワークを生成したり、分子機能ネットワークに伴う新規なイベント情報を取得したりできる。これにより、疾患における新たな分子的メカニズムの推定が可能になる。利用者によって編集されるデータは予め提供されたデータベースのデータ項目に追加・置換する形で記録してもよいし、編集用のファイル又はデータベースとして独立して管理し、必要に応じて予め提供されたデータベース内のデータと併せて利用してもよい。

【0101】

本発明のデータベース以外の外部データベースに収められた生体分子、生体分子対、生体イベント、病態イベント、医薬分子情報、その他の情報の全体または一部を、本発明のデータベースに取り込むことも可能である。

【0102】

本発明のさらに別の態様として、端末で閲覧中の情報に本発明のデータベース中のいずれかのデータ項目や生成した分子機能ネットワーク上のいずれかの項目に

該当する項目が存在した場合に、該項目を強調して表示する方法が提供される。この方法により、該項目が既にデータベース中に登録されていることを、わざわざデータベースに対して検索することなく利用者が知ることが可能になる。また、この際に該項目を強調表示だけでなく関連するデータベースのデータ項目や分子機能ネットワークの一部又は全部のデータを端末に表示することも可能である。さらには、該項目に基づいてコネクト検索を行い、該項目が関わる分子機能ネットワークを生成して表示することもできる。

【0103】

本発明の実施のさらに別の態様として、本発明の方法を実施するためのプログラムとデータベースからなるコンピュータシステム；本発明の方法を実施するためのプログラムとデータベースを記録したコンピュータ読み取り可能な媒体；本発明の方法で用いるためのデータベースを記録したコンピュータ読み取り可能な媒体；本発明の方法により生成された分子機能ネットワークに関する情報を記録したコンピュータ読み取り可能な媒体なども提供される。

【0104】

本発明の方法の特徴をまとめると、以下の通りである。

- ・ 生体イベントの情報を含み、直接結合する生体分子対の情報を蓄積することにより、生体内の分子間の関わりのデータベースを作成する。
- ・ パーツの集合である上記データベースからコネクト検索で、任意の生体分子又は生体イベントに関わる分子機能ネットワークを作成する。
- ・ 分子機能ネットワークに基づいて、任意の分子が直接間接に関わる生体イベントを推定する。
- ・ 生体イベントの情報つき分子機能ネットワークから、疾患メカニズム、可能な創薬ターゲット、副作用リスク等を推定する。
- ・ 生体分子の量的又は質的な変動から生体イベントのアップオアダウンを推定する。
- ・ 生体分子の生成臓器、存在臓器及び作用臓器の情報をもつ分子機能ネットワーク
- ・ 医薬分子情報と分子機能ネットワークを用いた副作用、薬剤間相互作用の推定

- ・医薬分子の投与による蛋白発現の変化を分子機能ネットワーク上で解釈
- ・遺伝情報とのリンクによって分子機能ネットワークへの遺伝子多型の影響、疾患背景遺伝子等の解析を行う。

【0105】

【実施例】

以下、本発明を実施例によりさらに具体的に説明するが、本発明の範囲はこれにより限定されるものではない。PCT/JPO1/07830号明細書に記載された実施例を参照することにより、下記の実施例の理解がさらに容易になる。

【0106】

例1

2以上の生体分子が複合体（集合体）として機能している場合の、生体分子情報データベースと生体分子連鎖データベースへの該複合体の登録方法と利用方法について、転写因子NF κ Bの系とTNFレセプター複合体を例にして示す（図5）。

転写因子であるNF κ Bは、RelAとp50のヘテロ2量体として機能する。このような分子については、以下のようにして1つの分子として扱うことが可能である。生体分子情報データベース中に、分子アノテーションの1項目として「複合体情報」の欄を設け、NF κ Bを構成する分子がRelAとp50であることを入力する。リガンド依存的に形成される複合体、例えばTNFレセプター複合体（図5中のTNFR1[4]）についても、同様に1つの分子として扱うことにより、分子機能ネットワークを簡潔に表現することが可能となる。

【0107】

生体分子によっては、あるときには複合体の構成分子として扱われたり、別のときには単独で1つの生体分子として扱われたりと、両方の場合を考慮する必要がある。図6に示すように、生体分子TRAF2は、複合体TNFR1[4]の構成分子として扱われたり、NF κ Bの下流で発現を誘導される単独の生体分子として扱われたりすることがある。この場合、独立な生体分子としてのTRAF2と、複合体TNFR1[4]の構成分子としてのTRAF2の間を関連付けすることにより、分子ネットワークをより緻密に表現することが可能となる。この関連付けは、分子情報データベース

中の分子略号と複合体情報欄にある分子略号とを照合することにより、自動的に
行なうことができる。

【0108】

例2

生体分子の修飾状態を生体分子データの階層化により区別する例を、NF κ B/
I κ B α 複合体の系について示す(図7)。
NF κ B/I κ B α は、まずIKK complexによってリン酸化され、次に(リン酸
化部位を認識して結合する)SCF complexによってユビキチン化され、最後に26S
proteasomeによってI κ B α が分解される。この例では、未修飾状態のNF κ
B/I κ B α 、リン酸化されたNF κ B/I κ B α 、さらにユビキチン化された
NF κ B/I κ B α を、生体分子情報データベースの「修飾状態」の階層のデー
タにより区別して扱うことができる。このように階層化データを用いることによ
り、NF κ B/I κ B α の各修飾状態に対して異なる分子略号を割り振ることな
くIKK complex等の各生体分子とNF κ B/I κ B α との関係を簡潔に表現する
ことが可能となる。

【0109】

例3

生体分子の修飾状態に関する情報が部分的にしか分かっておらず、生体分子対の
形成や生体イベントへの関連付けにおいて明確に区別できない場合も多い。例え
ば、図9に示す系では、異なる部位がリン酸化された生体分子p53と、生体分
子対を形成する相手の生体分子との対応関係が明確ではない。このような場合に
は、これらの複数の修飾状態にある生体分子を一つにまとめて扱ってもよい。

【0110】

例4

糖尿病を注目疾患として、病態連鎖データベースに含まれる各データ項目と、そ
れらの間の関連付けのデータの例を示す。

糖尿病は、1997年ADA (American Diabetics Association) により、1型、2
型、その他の特殊型、妊娠糖尿病と分類されているが、一般には遺伝因子と環境
因子が複雑に絡み合った、多因子疾患として捉えられている。

病態連鎖データベース中に含まれるデータ項目の例として、表1に糖尿病を注目疾患として本件発明の方法により記述した病態連鎖データベースの内容を示す。図9には、各データ項目の関連性を概念的に示した。図中で線で結ばれたデータ項目の対を、それぞれ関連付け情報として病態連鎖データベースに記録する。

【0111】

【表1】

キイ分子	生体反応	症状・症候
adiponectin(-)	beta3アドレナリン受容体遺伝子変異	beta細胞機能障害
b3-adR(-)	beta3アドレナリン受容体機能低下	アシドーシス
b3-adR(m)	DRB1遺伝子座との相関	アセトシス
FFA(+)	GIP機能障害	インスリン抵抗性
GIP(+)	GIPに対するインスリン応答減弱	家族性高プロインスリン血症
Glck(-)	GIP分泌能亢進	化膿傾向
Glck(m)	GLUT4移行抑制	かゆみ
D-Glc(-)	GLUT4発現抑制	クースマル呼吸
D-Glc(+)	HLA class II(DQA1)との相関	空腹時低血糖
HNF-1a(m)	HLA class II(DRB1)との相関	クッシング症候群
HNF-1b(m)	HNF-1alpha遺伝子変異	月経障害
HNF-4a(m)	HNF-1beta遺伝子変異	高PAI-1血症
insulin(-)	HNF-4alpha遺伝子変異	高レプチン血症
insulin(+)	IPF1遺伝子変異	神経痛

臨床検査値	疾患・合併症
75g経口ブドウ糖負荷試験(OGTT)2時間値(mg/dl) ≥ 200	1型糖尿病
Base excess(nmol/l) > -2	2型糖尿病
beta2ミクログロブリン(micro g/day) ≥ 5000	MODY1
betaカルテノイド(カルテノイド分画)高値	MODY2
HbA1C(%) ≥ 6.5	MODY3
HDLコレステロール(mg/dl)20-35	MODY4
HLA-DR3(白人)	MODY5
HLA-DR4(日本人)	Rabson-Mendenhall症候群
HLA-DR4(白人)	インスリン受容体異常症B型
ICA抗体(+)	インスリン受容体異常症A型
Val濃度(血漿)上昇	インスリンノーマ
アポリポ蛋白B1 ≥ 141	高LDL血症
アポリポ蛋白CII ≥ 7.7	持続性高インスリン性新生児低血糖症(PHHI)
アポリポ蛋白CIII ≥ 19.2	

【0112】

糖尿病のように多数の因子が関与している疾患では、メカニズムを明確な形で図式化し理解することは非常に困難であるが、本発明の方法を用いると、病態連鎖データベースの各グループに分類された個々のデータ項目がどのように関連しているか、また複数のデータ項目間の関連付けがどのように絡み合っているかを、容易に理解することができる。複雑な疾患の全体像、具体像を捉えようとするとき、疾患をその疾患に関与する病態イベント、生体分子、生体反応などの間の関連性という観点から理解することが可能となる。

【0113】

例5

病態連鎖データベース中の糖尿病に関与しているキイ分子から分子機能ネットワーク中のサブネットを検索する例を示す。

現在、糖尿病発症のメカニズムの詳細は明らかになっていない。そこで、まず、糖尿病において量的かつ／または質的に変動する分子を病態連鎖データベース内で検索し、TNF α (+)、FFA(+)、insulin(-)、HNF-4a(m)、glucose(+)、leptin(+)、HbA1C(+)という記述があるキイ分子群を得る。次に、生体分子情報データベースと生体分子連鎖データベースを用いて、該キイ分子のそれぞれが所属するサブネットを検索する。

【0114】

上記の手順により、TNF α (+)では「TNFレセプターを介したシグナル伝達」と「NF- κ Bによる発現誘導」、FFA(+)=「リポタンパク代謝」、insulin(-)では「インスリンシグナル伝達」と「グルコースによるインスリン分泌」、HNF-4a(m)は「膵の発生と代謝に関与する転写因子ネットワーク」、glucose(+)=「グルコースによるインスリン分泌」、「グリコーゲン代謝」と「解糖系」というように、それぞれのキイ分子に対して糖尿病に関与する可能性があるサブネット群が抽出される(図10)。

この方法により、糖尿病という疾患に関与する、あるいはこの疾患において変動する分子が、生命現象におけるどのようなパスウェイに関わっているかを調べることができる。

【0115】

例 6

糖尿病では現在、成因の違いによる分類法が用いられている。これは糖尿病の最近の成因の研究の進歩を踏まえ、より適切な診断と治療を行うことを目的として用いられている。ここでは、成因により分類された糖尿病のメカニズムを解析する例を示す。

「その他の特殊型」として分類される糖尿病の中で、MODY1(maturity onset diabetes of the youth1)が知られている。このMODY1が他の分類の糖尿病と比較してどのような違いがあるかを調べるために、病態連鎖データベース中の疾患・合併症のグループからMODY1の項目を検索し、該項目から病態連鎖データベースを用いてコネクト検索を行い、関連する項目を抽出する。

【0116】

上記の検索の結果、キイ分子のグループにMODY1の原因遺伝子に対応するHNF-4 α (m)が、生体反応のグループに「HNF-4 α 遺伝子変異」が、それぞれ見出される。さらに、生体分子連鎖データベースを用いて、転写因子であるHNF-4 α が関連付けられているサブネットを検索し、サブネット「脾の発生と代謝に関与する転写因子ネットワーク」を得る。該サブネットに属する分子機能ネットワークを生成することにより、実際にHNF-4 α 周辺の具体的な生体分子対も知ることができる(図11)。

このように、本発明の方法を用いると、より細かく分類された疾患名からでも、その成因に関わる分子および、メカニズム、あるいは他の疾患との差異に関する情報を得る事ができる。

【0117】

例 7

糖尿病が代謝異常による疾患であることは言うまでもないが、生体における代謝系と糖尿病の因果関係が直接明らかになっている疾患にMODY2が挙げられる。ここでは、糖質代謝と糖尿病、さらにその成因を目的とした検索例を示す。

まず、MODY2を質問として病態連鎖データベースを検索して、MODY2に関連するキイ分子のデータ項目GlcK(m)とGlcK(-)を得る。次に、GlcK(m)とGlcK(-)を質問として、生体分子連鎖データベースを用いて検索を行い、これらの質問項目に対応

するデータを持つサブネット「解糖系」を抽出する（図 1 2）。

【0 1 1 8】

例 8

糖尿病における最も重要な分子のひとつインスリンの生体における役割を調べるために、インスリンを質問として本件発明のデータベースを用いた検索を試みた例を示す。

糖尿病では何らかの形でインスリンが関与している。まず、インスリンを質問として病態連鎖データベースを検索し、キイ分子のグループからinsulin(+), insulin(-)及びinsulin(m)を、生体反応のグループから「インスリン作用低下」、「インスリン分泌能低下」、「インスリン受容体数減少」などを、症状・症候のグループから「インスリン抵抗性」を、臨床検査値のグループから「インスリン抗体(IAA)(+)、インスリン注射歴(-)」を、疾患・合併症のグループから「インスリン受容体異常症A型」などの項目を、それぞれ抽出する。

【0 1 1 9】

さらに、インスリンを質問として生体分子連鎖データベースを用いた検索を行い、サブネット名として「インスリンシグナル伝達」、「膵β細胞におけるインスリン分泌」などを抽出する。そのうち、特に「インスリンシグナル伝達」では、インスリンーインスリンレセプターの生体分子対の下流に「細胞増殖」、「蛋白質合成」、「グリコーゲン合成」、「脂質分解」、「解糖系」、「糖輸送」、「アポトーシス」などの生体イベントが存在していることがわかる（図 1 3）。したがって、このシグナル伝達は何らかの原因によって障害された場合、その影響も多岐にわたることがわかり、インスリンの生体における重要性和機能の多様性を知ることができる。

【0 1 2 0】

例 9

生活習慣病は幾つかの疾患が同時にまたは、付随的に発生し、共通の原因を持つことが多い。糖尿病それ自体も種々の合併症を引き起こすが、いわゆる「死の四重奏」や「シンドロームX」と呼ばれる症候群では、糖尿病だけを一つの疾患として切り離して考えることは困難である。

そこで、これらの症候群に共通する病態「インスリン抵抗性」を質問として病態連鎖データベースをコネクト検索してみた。コネクト検索の結果を有向グラフとして表示することにより、「インスリン抵抗性」が関与する疾患とその関連性を理解しやすい形で図示することができた(図14)。

【0121】

例10

医薬分子情報データベースのデータに基づいて、ある医薬分子の作用メカニズムを知る方法の例を示す。

医薬分子情報データベース中に含まれる情報の例として、糖吸収遅延を作用機序とし、糖尿病薬として用いられているアカルボースの医薬分子情報データベース中のデータを表2に示す。

【0122】

【表 2】

医薬分子情報データベース

医薬分子名 (製品名)	アカルボース- α -グルコシダーゼ阻害剤
ターゲット分子	α -グルコシダーゼ
副作用情報	<p>(1) 重要な副作用 劇症肝炎等の重篤な肝障害があらわれることがある。 肝機能障害、黄疸、AST(GOT) ALT(GPT)の上昇を伴う重篤な肝機能障害、黄疸が現れることがある。</p> <p>(2) その他の副作用 浮腫(特に女性)、腫脹感、LDH上昇、頭痛、ふらつき、頭重感 腹部膨満感、下痢、腹痛、胃痛、胃重感、腹部違和感(消化器) GOT上昇、GPT上昇(肝臓) 貧血、白血球減少、血小板減少(血液) 発疹、蕁麻疹、痒感(過敏症) BUN上昇(腎臓) 2型糖尿病</p>
対応疾患	糖尿病の食後過血糖の改善(ただし、食事療法、運動療法によっても十分な血糖コントロールが得られない場合、または食事療法、運動療法に加えて経口血糖降下薬もしくはインスリン製剤を使用している患者で十分な血糖コントロールが得られない場合に限る)
コメント	

【0123】

アカルボースまたはその商品名を質問としてその作用メカニズムを知ろうとする場合、医薬分子情報データベースを検索することにより、標的生体分子(ターゲット分子)が α -グルコシダーゼであるという情報を得ることができる。さらに、生体分子連鎖データベースを α -グルコシダーゼを質問として検索し、この分

子が「多糖・オリゴ糖代謝」というサブネットの中に存在することがわかる。さらに該サブネット中の分子ネットワークを表示させると、 α -グルコシダーゼがスクロースからグルコース、マルトースからグルコースなどへの分解反応を触媒しており、アカルボースはこれらの反応を阻害していることがわかる（図15）。

【0124】

例11

2型糖尿病の典型的な特徴はインスリン抵抗性である。インスリン抵抗性を改善するとされる医薬分子の作用メカニズムの詳細を分子機能ネットワーク上で調べる方法の例を示す（図16）。

「インスリン抵抗性」を質問として病態連鎖データベース内で検索を行うと、まず、注目疾患を「糖尿病」とする症状・疾患のグループに、「インスリン抵抗性」が見出される。次に、同じく「インスリン抵抗性」を質問として医薬分子情報データベースを文字列部分一致の条件で検索すると、インスリン抵抗性改善を薬理作用にもつ医薬分子であるチアゾリジン系の「ピオグリタゾン」と「トログリタゾン」（現在日本では販売中止）が抽出される。医薬分子情報データベースのデータから、これらのインスリン抵抗性改善薬の作用機序は「糖取り込み増加」、作用メカニズムは「核内受容型転写因子PPAR γ (peroxisome proliferator activated receptor γ) の活性化」であることがわかる。

ここで、作用機序「糖取り込み増加」について、「糖取り込み」を質問として、さらに分子機能ネットワークの検索を行うと、サブネット「インスリンシグナル伝達」における分子GLUT4の下流の生体イベントに「糖取り込み」があることがわかる。

【0125】

【発明の効果】

生体イベントを含む生体分子対の情報の集合である本発明の生体分子連鎖データベースは、検索によって必要な範囲の機能上生合成上の該分子間のつながりである分子機能ネットワークを生成し、任意の生体分子が発現に直接又は間接に関わる生体イベントを推定するのに有用であり、さらに医薬分子の情報又は遺伝情報

とリンクすることにより、新薬開発や個体差に基づく医療に必要な知識を得ることが出来る。

【図面の簡単な説明】

【図 1】 本発明の方法の基本的な概念を表わす図である。

【図 2】 3 以上の端点を指定したコネクト検索の結果の例を示した図である。

【図 3】 本発明の方法で医薬分子連鎖データベースを用いる場合の概念を表わす図である。

【図 4】 本発明の方法で遺伝情報データベースを用いる場合の概念を表わす図である。

【図 5】 例 1 の複合体情報の生体分子情報データベースへの登録方法と利用方法を表わす図である。

【図 6】 例 1 の生体分子の複合体状態と単独状態の間を関連付けする方法を表わす図である。

【図 7】 例 2 の生体分子データの階層化により修飾状態を表現する方法を表わす図である。

【図 8】 例 3 の生体分子の異なる修飾状態を一つにまとめて扱う方法を表わす図である。

【図 9】 糖尿病を例にとった病態連鎖データベースのデータ項目の関連性の概念を表す図である。

【図 1 0】 例 5 の検索方法により得られた分子機能ネットワーク中のサブネットと病態連鎖データベース中のキイ分子との関連性を示す図である。

【図 1 1】 例 6 の疾患名や病態イベントから分子機能ネットワークを利用して疾患のメカニズムを調べる方法を示した図である。

【図 1 2】 例 7 における、疾患名からキイ分子を介してサブネットを検索する方法を表す図である。

【図 1 3】 例 8 の生体分子の生体における機能と役割を調べる方法を表す図である。

【図 1 4】 例 9 の病態連鎖データベースを用いた検索結果を表す図である。

【図 1 5】 例 1 0 の医薬分子情報データベース中の標的生体分子情報に基づい

て医薬分子の作用メカニズムを調べる方法を示す図である。

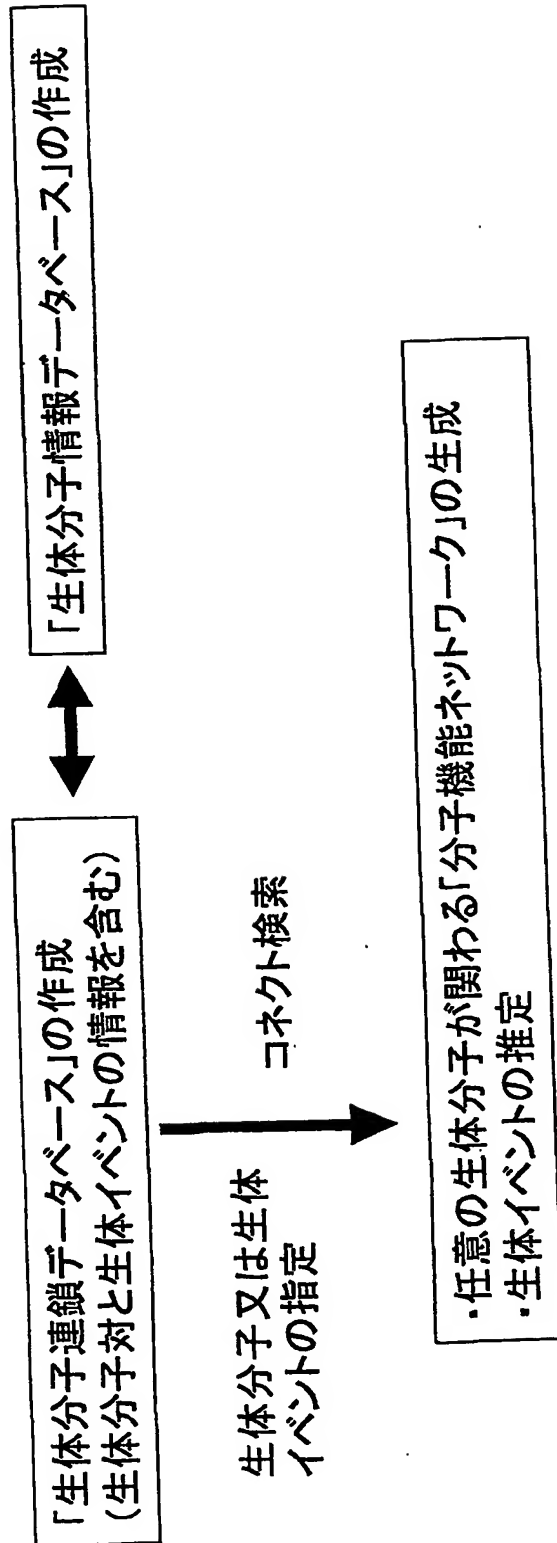
【図 1 6】 例 1 1 の生体分子、生体イベント、サブネット等を介して医薬分子の作用メカニズムの詳細を調べる方法を表す図である。

特 2 0 0 2 - 0 6 4 9 9 4

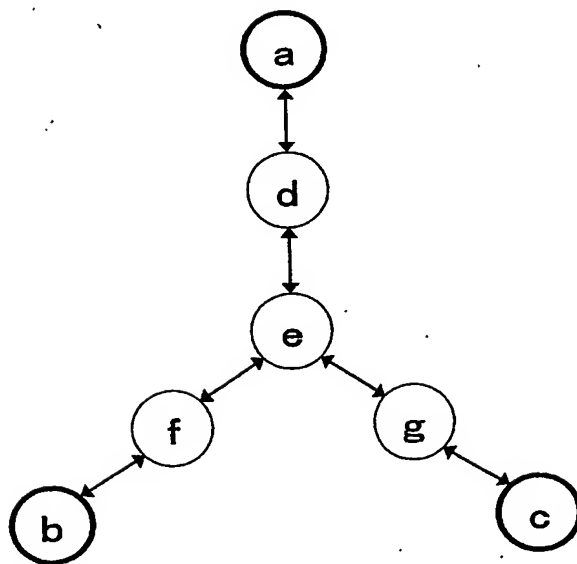
【書類名】

図面

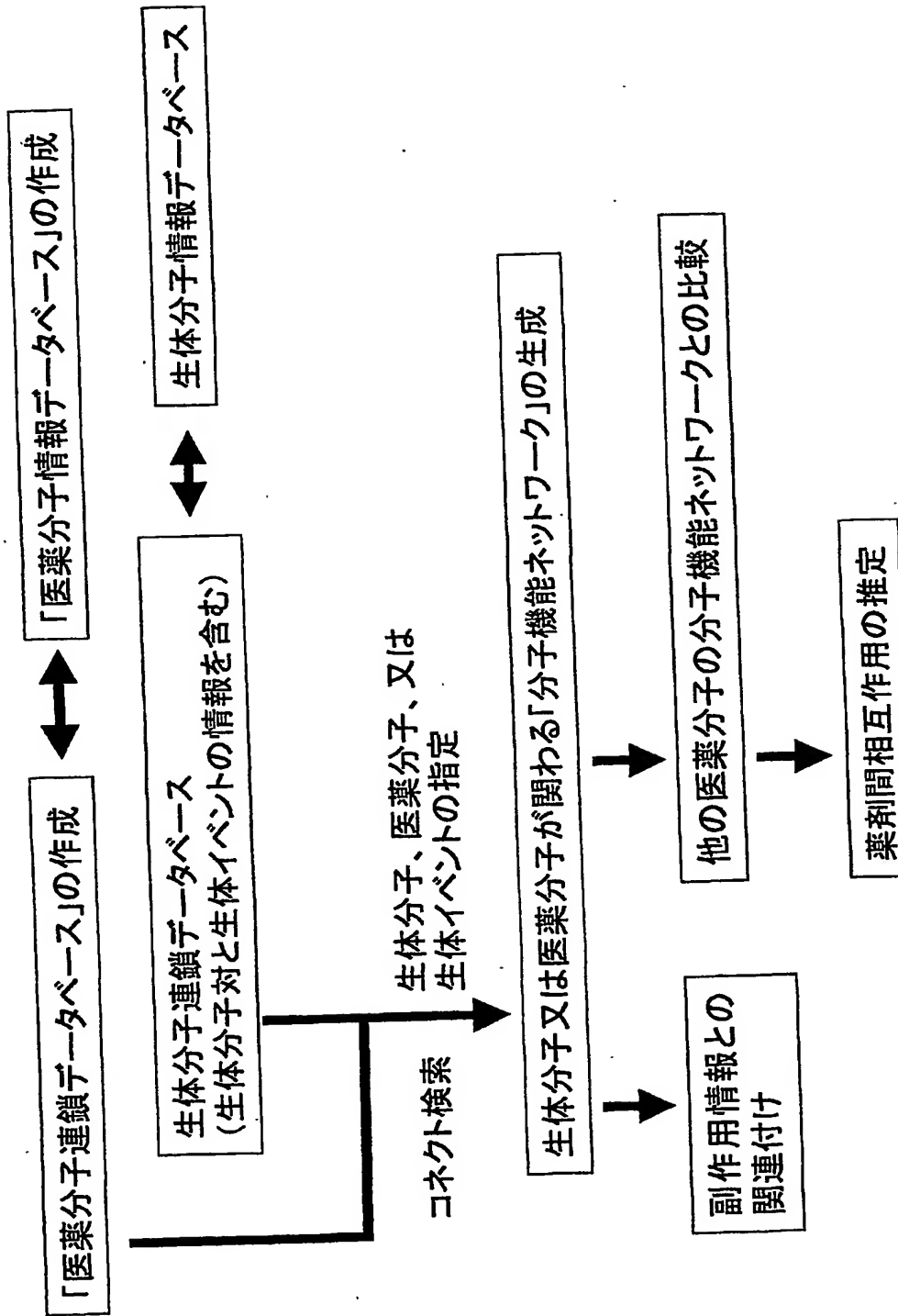
【図 1】



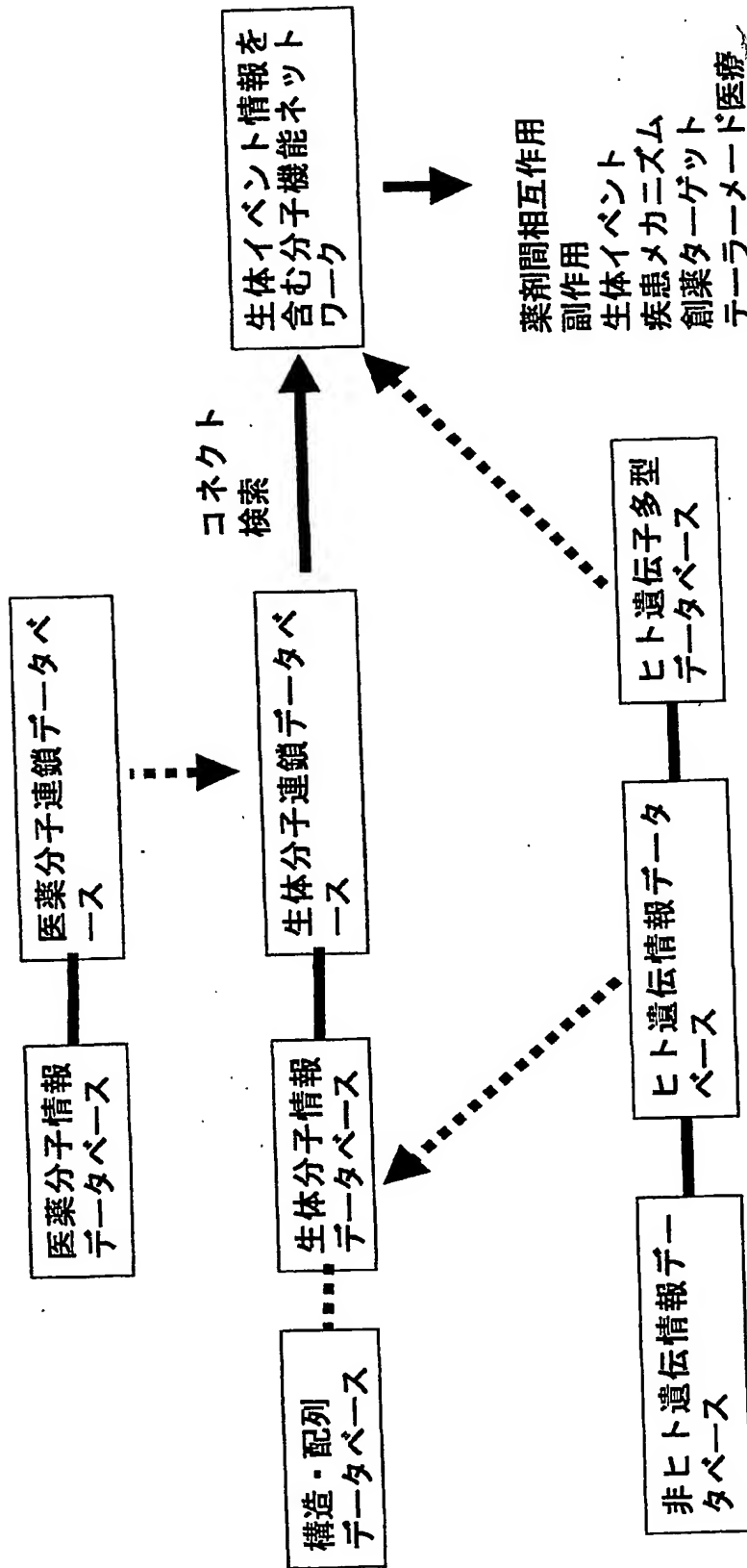
【図 2】



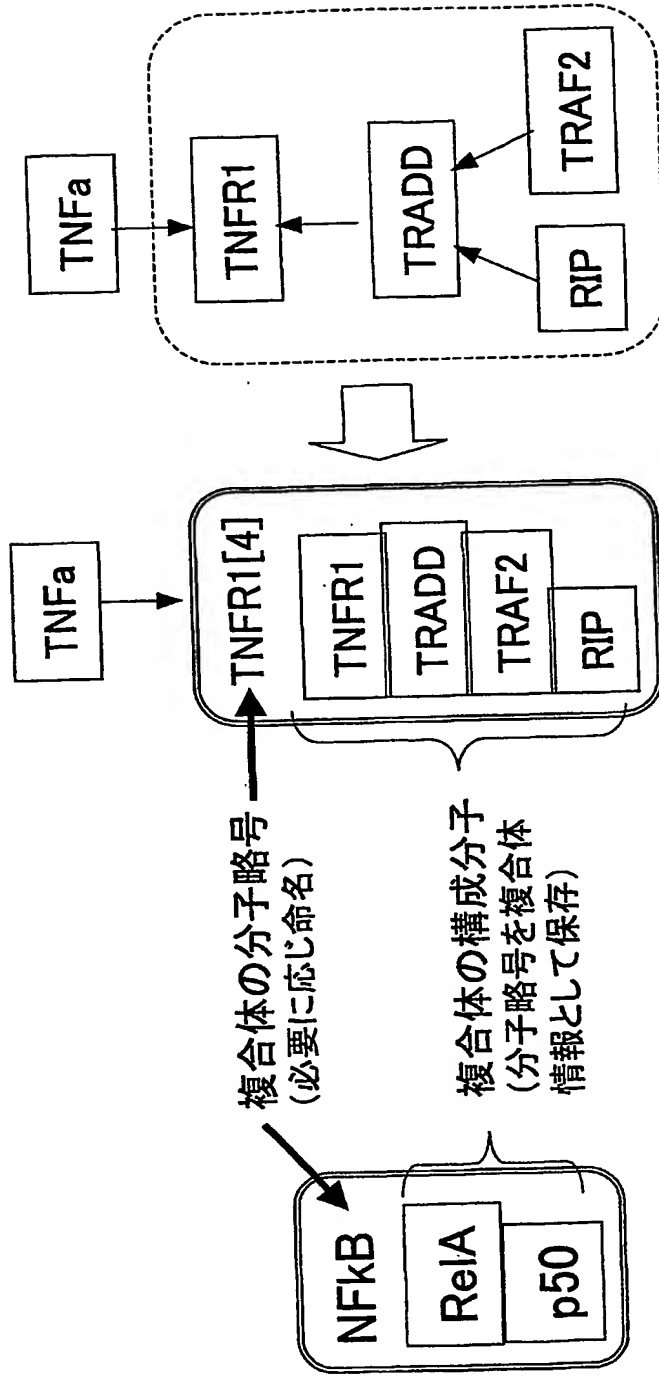
【図 3】



【図 4】



【図5】

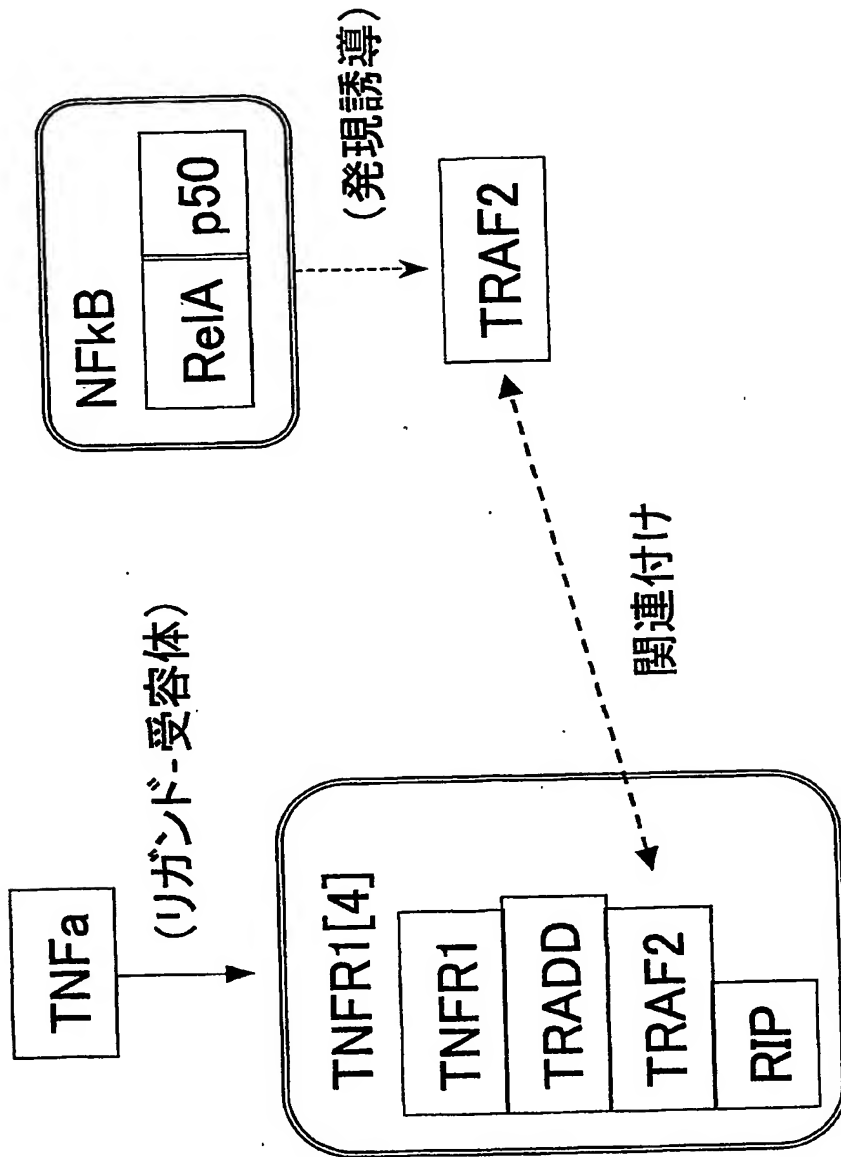


転写因子NFkB

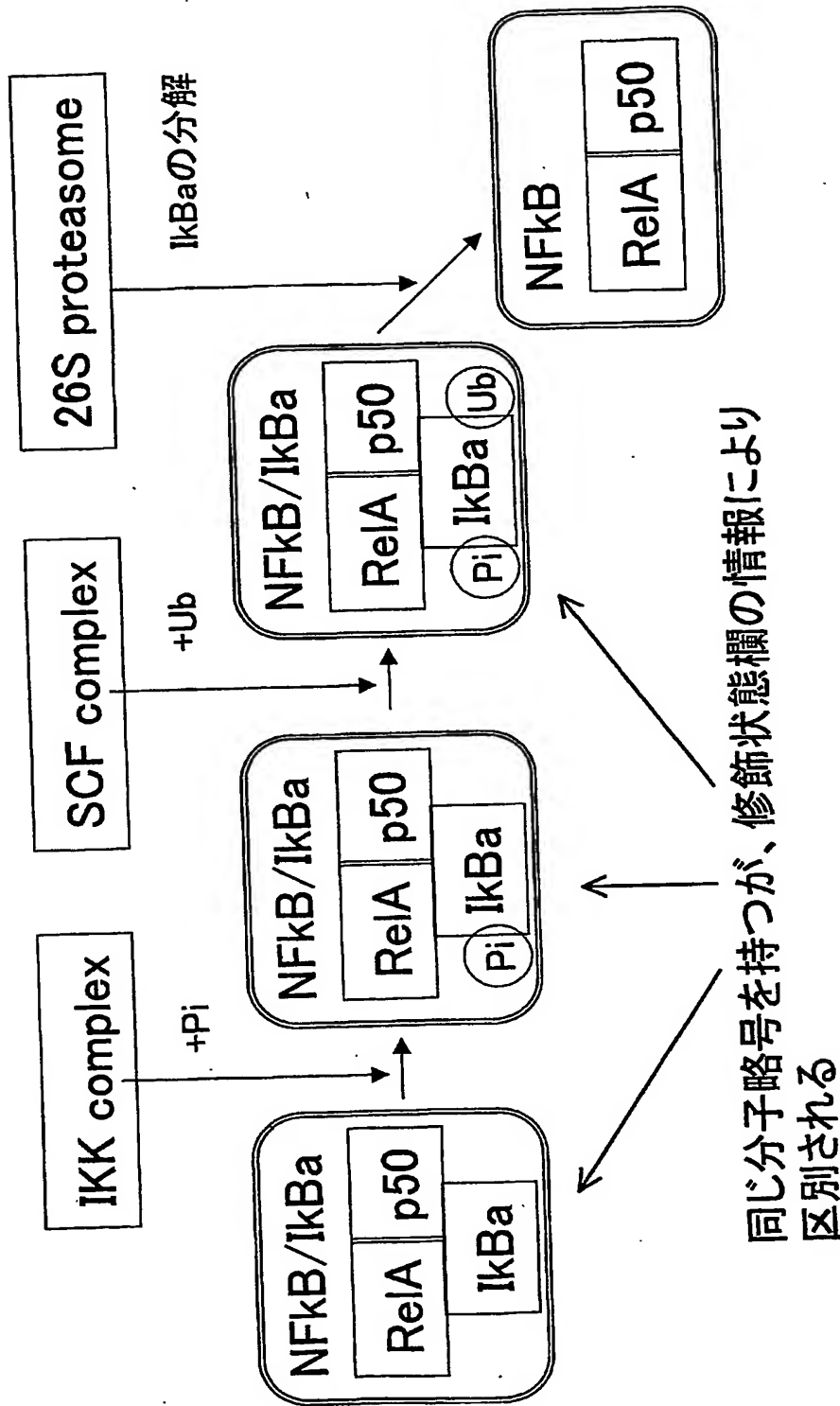
TNFa-TNFR1[4]
として生体分子対を記述

シグナル伝達の過程で
リガンド(TNFa)依存的
に複合体を形成

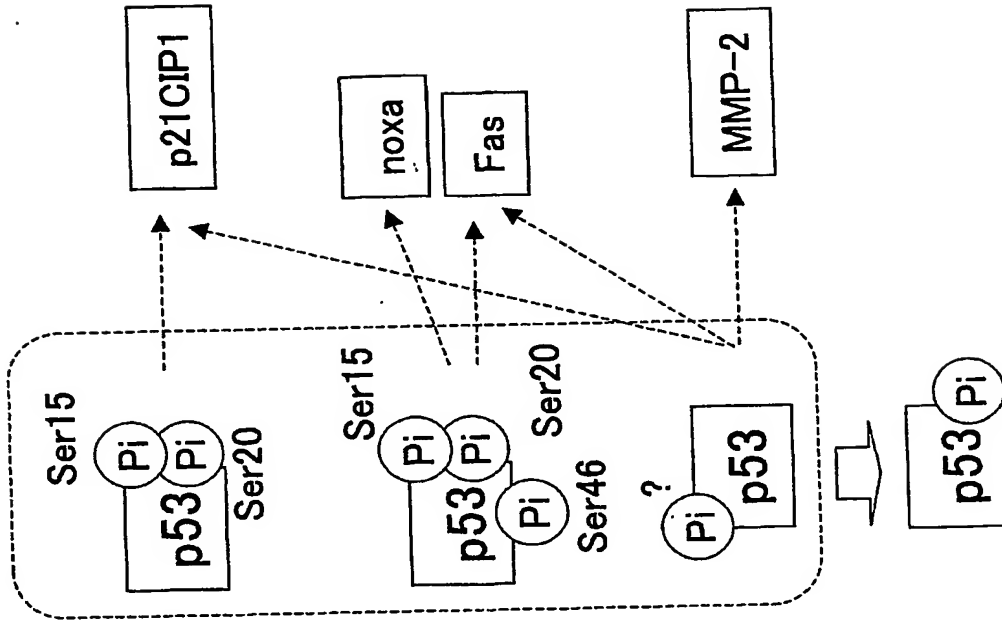
【図 6】



【図 7】

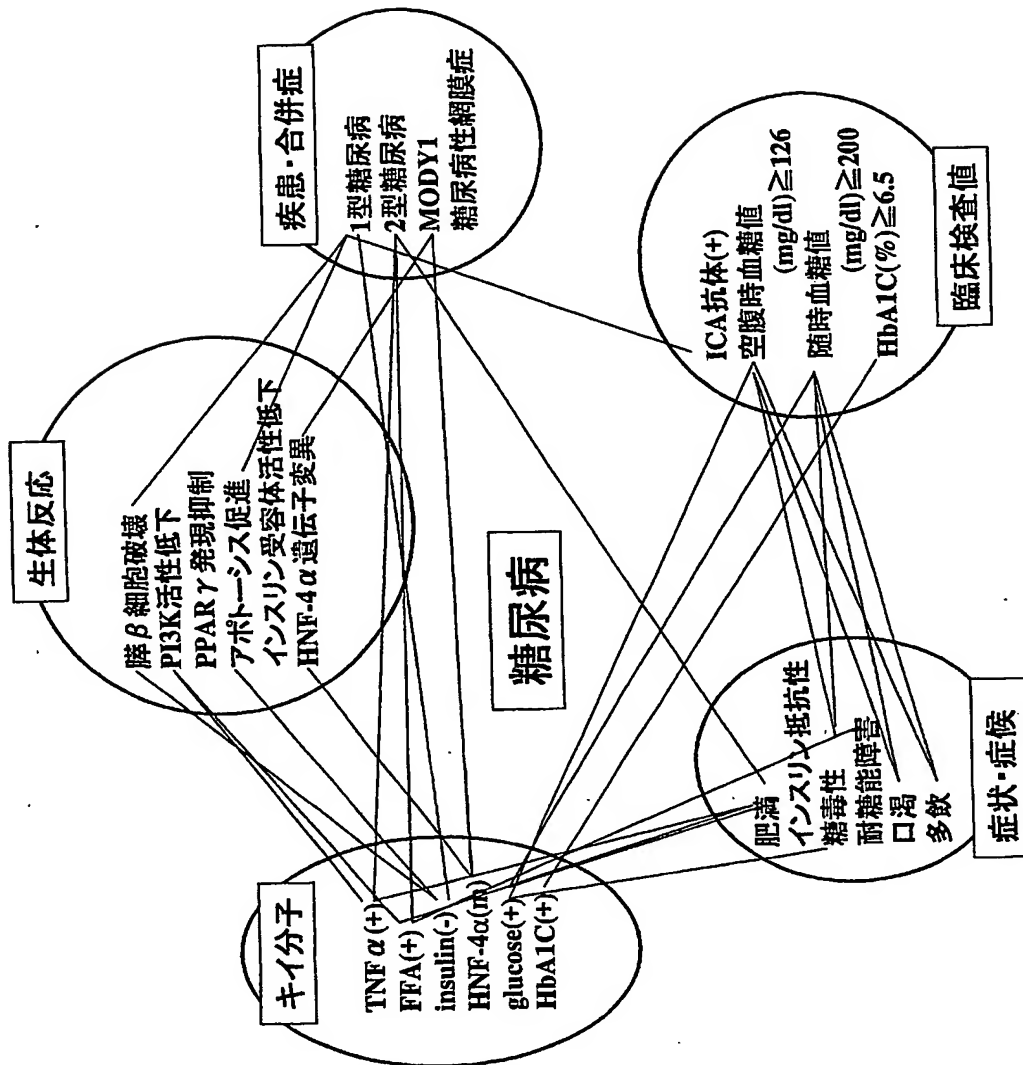


【図 8】

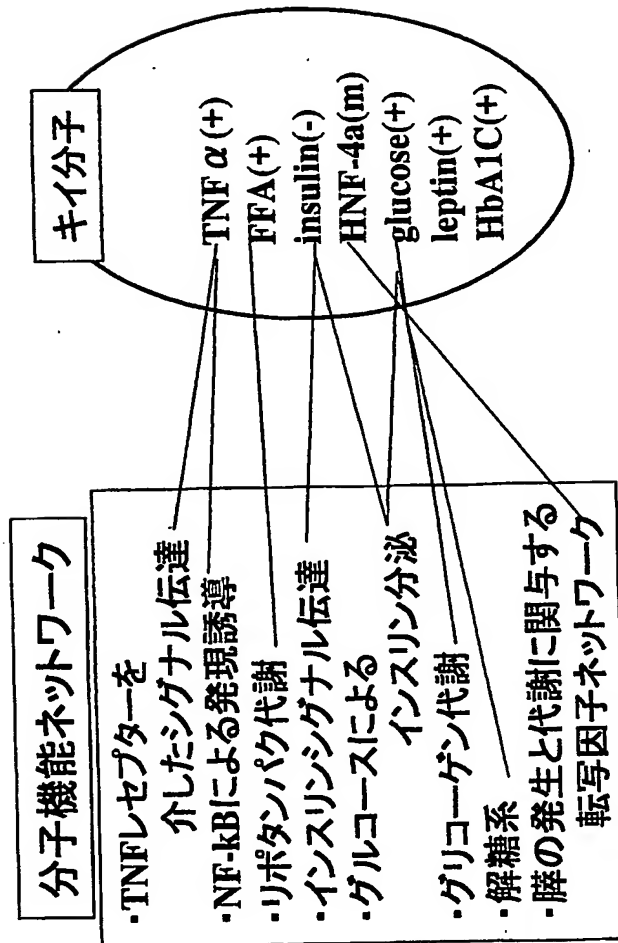


情報源により修飾情報が異なる
あるいは不足している
↓
矛盾なく生体分子対を形成できる
ように修飾情報を調整

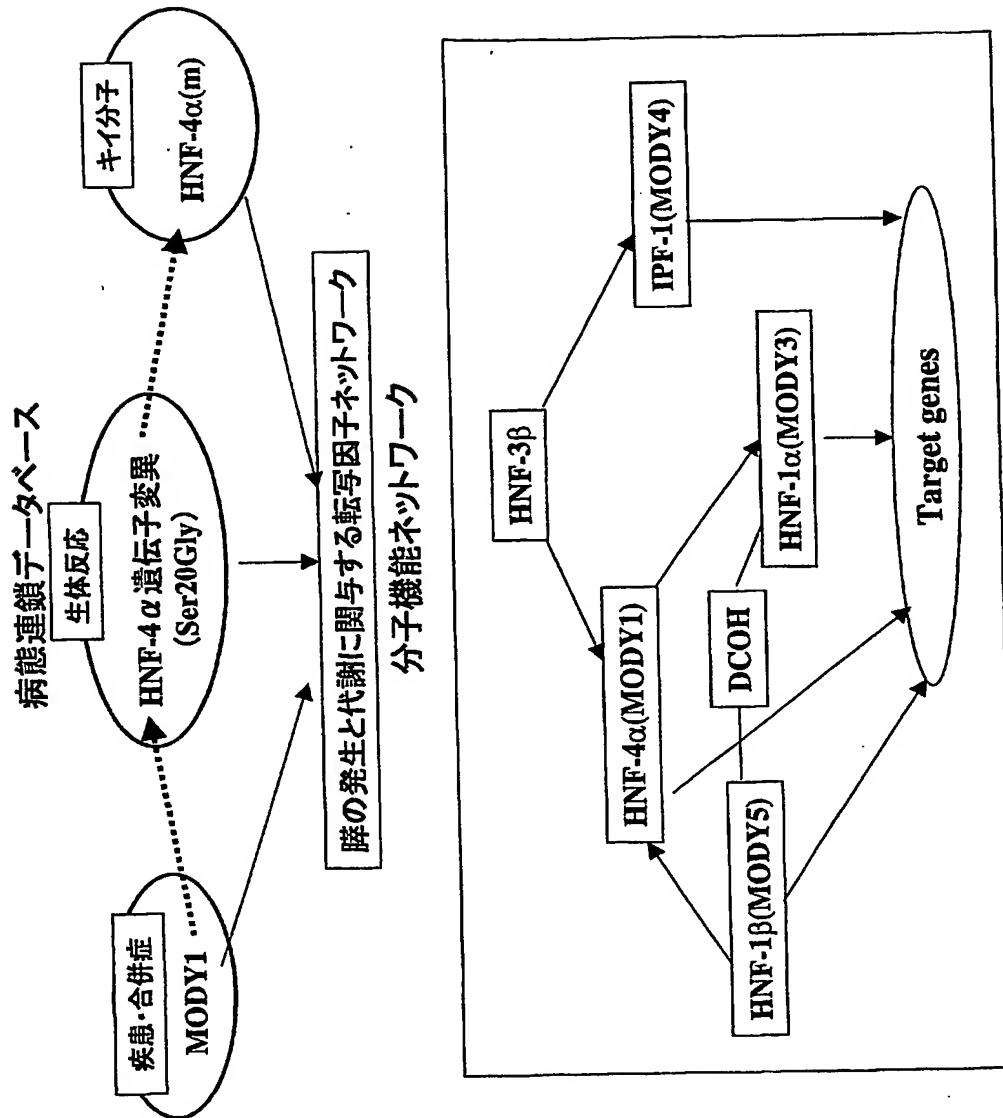
【図 9】



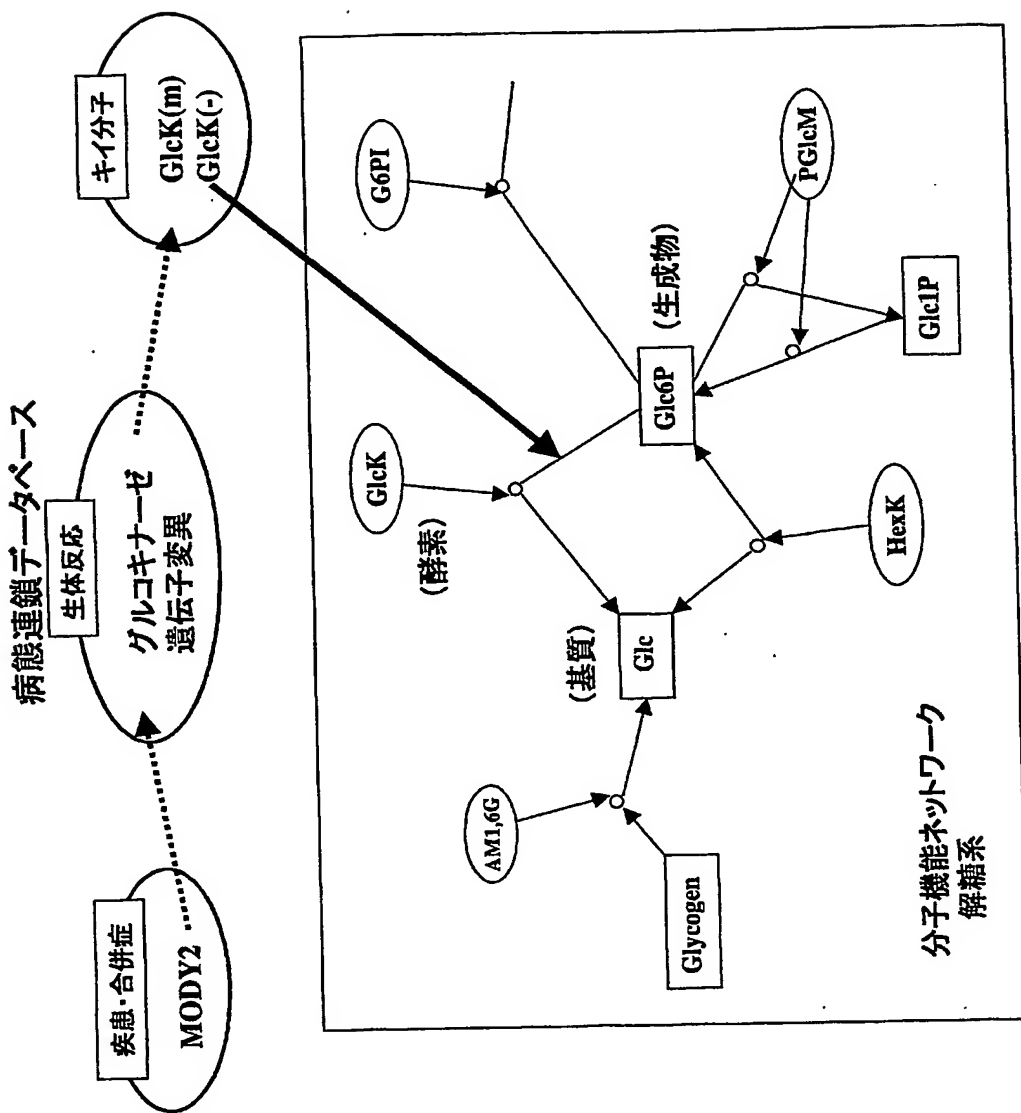
【図 10】



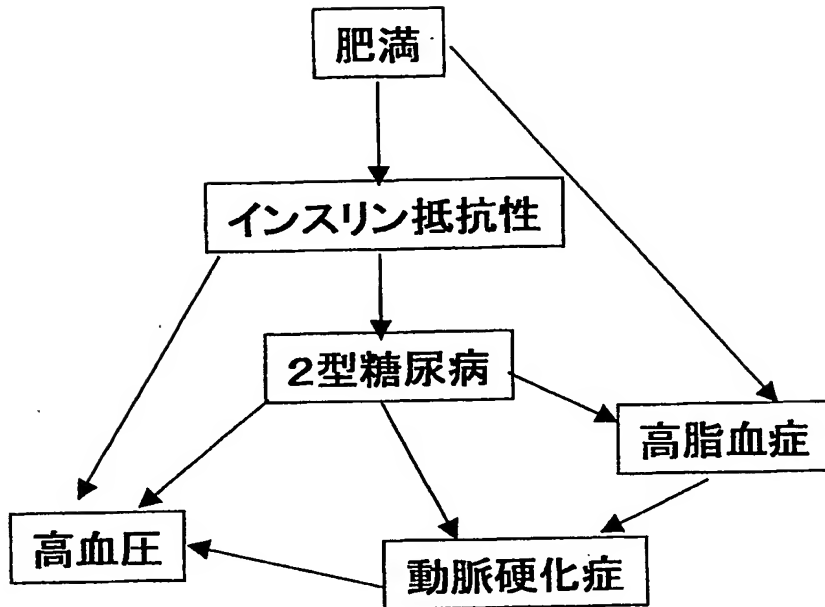
【図 1 1】



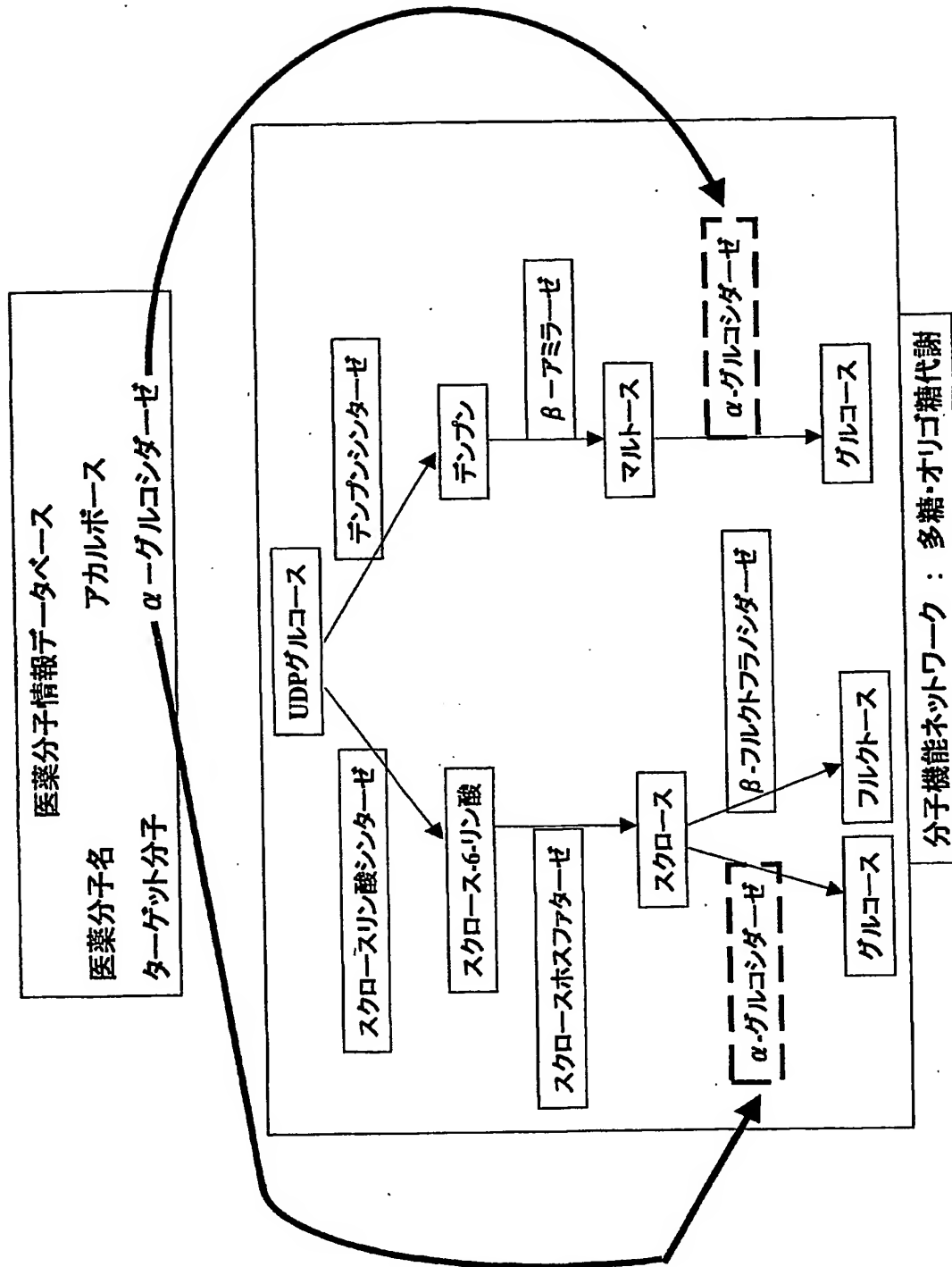
【図 12】



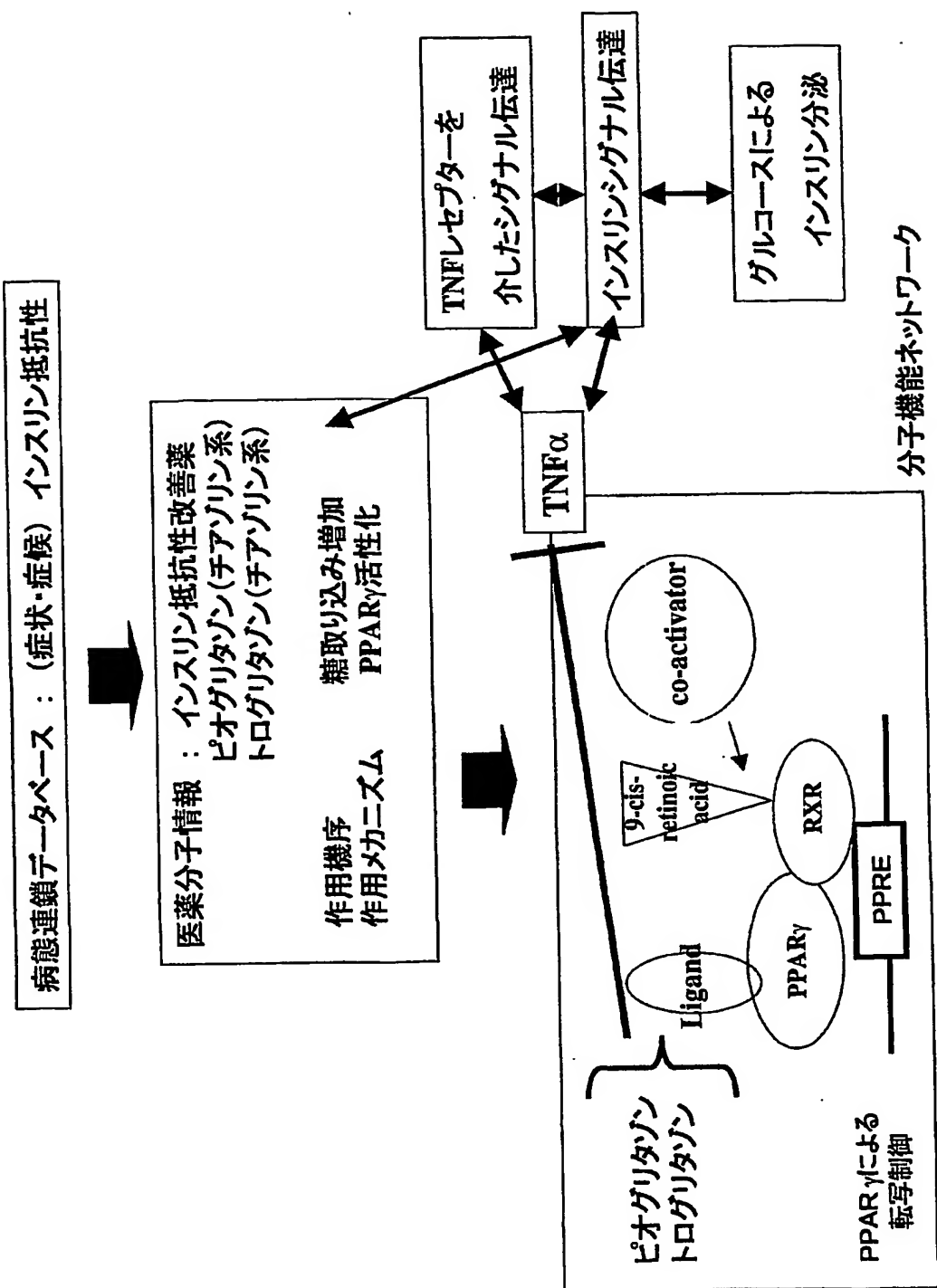
【図 14】



【図 15】



【図 16】



【書類名】 要約書

【要約】

【課題】 多様な生体応答や現象を生体分子の機能とその分子間の関わりにおいて理解するための手段として生体分子の情報を生体応答と結びつけることのできるデータベースとその検索法を提供する。

【解決手段】 生体分子の情報を修飾状態、活性／不活性状態、集合状態、構造変化を含む1又は2以上の項目により階層化して保存したデータベースを用いてコネクト検索を行うことを特徴とする、分子機能ネットワークの生成方法。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [597051148]

1. 変更年月日 1997年 4月11日

[変更理由] 新規登録

住 所 東京都文京区本郷5丁目24番5号 角川本郷ビル4F

氏 名 株式会社医薬分子設計研究所